

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 02:12:21 ; Search time 5404 Seconds  
(without alignments)  
11427.644 Million cell updates/sec

Title: US-09-932-678-1

Perfect score: 2068

Sequence: 1 acagagctgtgctggaag.....tgacattgggtcccatc 2068

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_estc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_estc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pin.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1474.2	71.3	3457	11 AK080397	AK080397 Mus muscu
2	1386.8	67.1	2136	11 AK080680	AK080680 Mus muscu
3	815.4	39.4	1800	14 YL6973	YL6973 RNY16973 Ra
4	711	34.4	896	13 BQ214727	BQ214727 AGENCOURT

5	675.6	32.7	894	14	CD359064
6	675.4	32.7	788	12	BG502871
7	670.8	32.4	682	14	CB150625
8	666	32.2	733	12	BI561523
9	653.6	31.6	996	10	BF203684
10	646	31.2	757	10	BE734354
11	637.4	30.8	1037	12	BG428305
12	628.6	30.4	889	13	BUS10623
13	627.2	30.3	1122	13	BQ066201
14	617.8	29.9	671	9	AV703279
15	614.6	29.7	736	12	BI464721
16	613	29.6	828	13	BU610641
17	588.4	28.5	645	13	BQ638571
18	582.2	28.2	990	12	BG167061
19	574	27.8	580	12	BM564978
20	571.4	27.6	753	14	CB266771
21	566.8	27.4	769	14	CF738084
22	566.6	27.4	2659	11	BC037959
23	566.4	27.4	568	9	AL708660
24	553.4	26.8	828	12	BI761160
25	550.4	26.6	742	10	BE870692
26	548	26.5	550	9	AL597171
27	547.8	26.5	551	12	BM786521
28	547.2	26.5	727	12	BI549394
29	542.4	26.2	563	10	AW408086
30	536.8	26.0	564	10	AW958173
31	532.6	25.8	651	14	CF908118
32	520.4	25.2	697	12	BG111364
33	517	25.0	535	9	AA213789
34	503.8	24.4	835	12	BI601238
35	500.8	24.2	764	13	EX101439
36	500.6	24.2	1065	13	BQ070706
37	498.6	24.1	609	14	CF902795
38	497	24.0	505	14	CF138931
39	493.6	23.9	712	13	BY733332
40	493.4	23.9	1011	13	BU151651
41	492	23.8	612	13	BQ771414
42	488.4	23.6	498	10	BF110176
43	487.4	23.6	497	10	BF221544
44	483	23.4	654	14	CF737683
45	482.8	23.3	498	12	BG149493

## ALIGNMENTS

RESULT 1  
AK080397  
LOCUS  
DEFINITION Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630092F02 product:RRN3 homolog [Homo sapiens], full insert sequence.  
ACCESSION AK080397  
VERSION AK080397.1 GI:26099204  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1  
AUTHORS Carninci,P. and Hayashizaki,Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itch,M., Komno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159

AK080397 3457 bp mRNA linear HTC 19-SEP-2003  
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630092F02 product:RRN3 homolog [Homo sapiens], full insert sequence.

REFERENCE AUTHORS	3	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Komono,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system -384-format Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	4	20530913 11076861 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
TITLE	5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II team. Analysis of the mouse transcriptome based on functional annotation of 6,770 full-length cDNAs Nature 420, 563-573 (2002)
JOURNAL REFERENCE AUTHORS	6	(bases 1 to 3457) Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoch,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physiological and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/ Location/Qualifiers
FEATURES	source	1. 3457 /organism="Mus musculus" /mol_type="mRNA" /strain="CS7BL/6J" /db_xref="FANTOM DB:A630092F02" /db_xref="MGI:2407489" /db_xref="taxon:10090" /clone="A630092F02" /tissue_type="thymus" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="3 days neonate" 14. 1987 /note="RN3 homolog [Homo sapiens] (SPTR Q9NVV6, evidence: FASTV, 87.8%ID, 92%length, match=1791) putative"
misc_feature		
ORIGIN		
Query Match	71.3%;	Score 1474.2; DB 11; Length 3457;
Best Local Similarity	85.3%;	Pred. No. 0;
Matches 1704;	Conservative	0; Mismatches 283; Indels 10; Gaps 5;
QY	63	GGCCCAATGGCGCCGCTGCTTACACAGCGTTCGCCGGGAGATCGCGCGTTCGTCC 122
DB	8	GGCCGAATGGCGCCGCGCTGCTTACACAGCGTTCGCCGGGAGATCGCGCGCGGCC 67
QY	123	TCCTGCAGTTAAGAAGCTGGCGCGCTCGAGACTGGGATTCAAATATATGCTGCTAGAG 182
DB	68	TCCTGCC-----ACGCTGAGTCGCTCGGACTGGGCTTCGATATGCTCGGTTAGAG 121
QY	183	AATGACTTTTCAATCTCTCCCAAGAAAAAAGTTCGTTGGTGGAACTGTGCACAGAA 242
DB	122	AGCGAATTTCTTCAATCTCTCCGAAAAAAAGTTCGTTGGTGGAACTGTGCACAGAA 181
QY	243	GTCTTGTGTAAGTACAAAAAGGTTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAGCTG 302
DB	182	GTCTTACTGAAATACAAAAAGGTTGAAACAAATGACTTTAGAGTTGTTGAAGAACCCAGCTG 241
QY	303	TTAGATCCAGACATAAAGGATGACCAAGATCATCAACTGGCTGCTAGAAATTCGGTTCTTCT 362
DB	242	TCCTGATCCTGATATTAAGATGACCAAGATCATTAAGTGGCTACTAGAAATTCGTTCTCT 301
QY	363	ATCATGTACTTGACAAAAAGACTTTGAGCAACTTATCAGTATTTATTAAGATTTGCTTGG 422
DB	302	GTCTGTGTTTTCGACAAAAAGACTTTGAGCAACTTATTAAGATTTGAGATTTGCTTGG 361
QY	423	TTGAATAGAAGTCAACACAGTACTGGAAGAGTATTTGGCTTTTCTTGGTAAATCTTGTATCA 482
DB	362	TTGAATAGAAGTCAAGAGAGTGGTGGAGAGTATTTGGCTTTTCTTGGTAAATCTTGTGTCT 421
QY	483	GCACAGACTGTTTCTCCTCAGACCGGTCTCAGCATGATTCCTCCCAATTTTGGCTCC 542
DB	422	GCACAGACTGTTCTCTTTAGACCAATGTTCTCAGCATGATTCGTTCTCAATTTTGTACTCC 481
QY	543	CGAGTATCATTAAGGAAGGGATGTAGATGTTTCAGATTCGTATGATGAAGATGATAAT 602
DB	482	CGAGTAAATTTGTCAAGGAAGGTGGCATAGATGTTTCAGATTCGTATGATGAAGATGACAT 541
QY	603	CTTCTCTGCAAAATTTTGACACATGTCACAGAGCTTGCAAAATTAATAGCAGATATGTACCA 562
DB	542	CTTCTCTGCAAAATTTTGACACATGTCACAGAGCTTGCAAAATTAATAGCAGATATGTACCA 601
QY	663	TCGACACCGTGTCTTCATGCCAATCTGTGGAAAAAATTTCCCAATTTTGTTCGAAAAATCA 722
DB	602	TCGACACCATGTTTCTTAATGCCAATCTGTGGAAAAAATTTCCCGTTTGTGAGGAAGTCC 661
QY	723	GAGAGAACTGGAAATGTAGCTTCATTAACCTTAAGAGATTAAGTATATTTTCCAAACC 782
DB	662	GAGAGAACTGGAAATGTATGTTTCAATTAACCTTAAGAGATTAAGTATATTTTCCCAACT 721
QY	783	TTGAGGCATGAAATTCCTGAGCTTATTATGAAGAACTACTCAAGTTCGTA-TGTGAATGC 841
DB	722	TTGAGGCATGAAATTCCTGAGCTTATTATGAAGAACTACTCAAGTTCGTA-TGTGAGTGT 781
QY	842	ATCCCGGAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGACAGA 901
DB	782	ATCCCGGAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAGACTTGTGGTGGACAGA 841
QY	902	TTCCACGGAAGGATTTGTTTAAATATGGATGAAGTGAAGAAACTGAAATGAACAAAGGC 961
DB	842	TACCACGAAGGACTGTTTAAATATGGATGAAGTGAAGAACTGAACTCCCTCGAAGAAAGC 901
QY	962	TGCTCTCGAAGCGGTCCGACCATGCTGTCATCTCTGTAGCCGAGCCCTGGACATCTCGAT 1021
DB	902	TGACGAGGAGCAGCTCAACAGATGGCTCATCCCAACCGCAGAGCGCTGGAGCTCTGCT 961
QY	1022	GTCTTTGTTTGTCTTACATGAAGGATGTCTGCTATGTAGATGTAAGTTGATAACGG 1081
DB	962	GTGCTTGCTGCTCTCAATTTGAGGATGTCTGCGCGTGTGACCGGTAATAATGATAACAA 1021

QY 1082 CRAAACAAAGAGTCTATATCGGACCTGATAACATCTTTGACAAACTCCTGTGTGCCAC 1141  
Db 1022 TAAACAAGAGATTATACCGTGATCTGATATCATCTTTTGACAAACTCTGTGTGCCAC 1081  
QY 1142 CCATGCTCTCGCATGTACAGTCTTTTCATGTTTTACCTCTGTAGTTTCAAATTTGGATT 1201  
Db 1082 ACATGCTCTCGCATGTACAGTCTTTTCATGTTTTCTCTGCGAGTCTCAAGTTGGATT 1141  
QY 1202 CGCAGAGGATTTTTTGGAAATCTCTGGAAAAATTTGCAGACCCAGTAATCCTGCCAT 1261  
Db 1142 TGCAGAGGATTTTTTGGAAATCTCTGGAAAAATTTGCAGATCCAAATAACCCGCCAT 1201  
QY 1262 CATCAGGAGGCTGTGGAAATTTATATTGAAGCTTTTTCGGCAAGAGCTAAATTTATTC 1321  
Db 1202 CATCAGGAGGCTGTGGAAATTTATATTGGAGCTTTTTCGGCCAGAGCTAAATTTATTC 1261  
QY 1322 TCTTATTACTGTAAATCATGCCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAA 1381  
Db 1262 TCTTATCAGCTGTGAAGACATGCTGGATCTCTTGGTTAACTGGCTGCACATATACCTTAA 1321  
QY 1382 TAACAGGATTCGGAAACAAAGGATCTCGGATGTTGCTCTCCATGGACCATTTTACTC 1441  
Db 1322 TAACAGGATTCGGAAACAAAGGCTTTTGTGACGTTGCACTCCATGGACCATTTTATTC 1381  
QY 1442 AGCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACAAAGCAGCTTTTGTAGCGGAAA 1501  
Db 1382 AGCTGCCAAGCTGTGTTCTACACTGTTGTTTGTAGACAAAGCAGCTTTTGTAGCGGAAA 1441  
QY 1502 CCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGTAGCGGATATGATGAGCCA 1561  
Db 1442 CTTGAAAGCAAGTCTACAGTATCTTCAGAGTCTGAATTTTGTAGCGGATATGATGAGCCA 1501  
QY 1562 GCTAAATCCCTGAAGATTTGCTGCGCTCAGTGGTTAACTTTTCTGTCGAATCAGAAA 1621  
Db 1502 GCTGAACCCACTGAAGATCTGCTGCGCAGAGTGGTTAAATTTCTGTCGTATCAGAAA 1561  
QY 1622 TAAGTACAGCTGCTCTCTGCTACACCATATTGAGAGGAACAATCGCCAGATGCTGCC 1681  
Db 1562 TAAATACAGCTGTTGTTCTGCTACACCATATGAGAGGAACAATCGCCAGATGCTGCC 1621  
QY 1682 AGTCATTAGGAGTACCGCTGGAGGAGCTCAGTGCAGATCTGCACAAACCCGCTGGACAC 1741  
Db 1622 CGTTATCGGACGACCGCTGGGAGACTCGGTGCAGACCTGCAACCAACCACTGGACAC 1681  
QY 1742 CTTCTCCCTTTGATCCCTGTGCTGAAGAGTCAAGAAATTCATTGATCTCTATTTA 1801  
Db 1682 TTTCTCCCTTTGACCCCTGTGCTTAAGAGTCAAGAAATTCATTGATCTCTATTTA 1741  
QY 1802 TCAGGTATGGGAAGA-CATGAGTGTGAAGAGCTACA-GGAGTTCAAGAAACCCATGAAA 1859  
Db 1742 TCAGATTGGGAAGATGGGAGTGTGAAGAGCTTCAGGGAGTTAAGAAATCTACTAAA 1801  
QY 1860 AAGGACATAGTGGAGATGAAGATGATGACTTTCTGAAGGCGAAGTCCCGCAGATGAT 1919  
Db 1802 AAGGAGGTAGTGGAGATGAAGATGATGACTTTTGAAGGCGAGGTCCCGCAGAGTGCAC 1861  
QY 1920 ACCGTGATGGATCAGACCAAGCTCCTTTGACCGCATTTCCGAGGCTCTCAAGTAGT 1979  
Db 1862 ACAGTGAAGTGGCTTACTCCGAGCTCCTTTGATATCCCACTTCCAAAGTCTTCCAGTAGT 1921  
QY 1980 GTGGGCTCCCAACCGTGTGTGTACAT-GCAACCCAGTCCCTCTGACGCGCAAAATTTGT 2038  
Db 1922 GTGGGCTCCCTCTCTGTGTGTATATACAGCGCAGTCTCCACTCTCACAAGGATCTAT 1981  
QY 2039 GACTGAGATGTGACATT 2055  
Db 1982 GATTGAGCTGCGCAATT 1998

RESULT 2  
AK080680  
LOCUS

## DEFINITION

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED

TITLE  
JOURNAL  
MEDLINE  
PUBMED

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL

## COMMENT

Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched library, clone:A830089A07 product:RKN3 homolog [Homo sapiens], full insert sequence.

AK080680  
AK080680.1 GI:26099414  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20493374  
11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

6 (Bases 1 to 2136)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsumi, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Tagabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216]

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.go.jp/>  
URL: <http://fantom.gsc.riken.go.jp/>.

## FEATURES

```

location/Qualifiers
1. .2136
/organism="Mus musculus"
/mol_type="mrna"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:A830089A07"
/db_xref="MGI:2409380"
/db_xref="taxon:10090"
/clone="A830089A07"
/tissue_type="cortex"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
40. .1905
/note="PBN3 homolog [Homo sapiens] (SPTB|Q9NYV6, evidence:
PASBY, 87.8%ID, 92%length, match=1791)
putative"
2110. .2115
/note="putative"
2136
/note="putative"

misc_feature
polyA_signal
polyA_site

```

## ORIGIN

Query Match	67.1%;	Score 1386.8;	DB 11;	Length 2136;
Best Local Similarity	84.5%;	Pred. No. 0;		
Matches 1608;	Conservative 0;	Mismatches 282;	Indels 12;	Gaps 4
QY	31	TCGGGCTCAGCGCAGCGGTGCGCTTAGTTTCGGCCCAATCGCGGCHCCGTGCTTCACA	90	
Db	2	TCGGGCGAAGCGGGGTTTCGCTCAGTTCAGGCCGAATGGCGGCCGCGTGCCTTCACA	61	
QY	91	CGGGTTTCGGGAGATCGGCGGCTTCGTCTCTGCGAGTTTAAAGACTGGCGGCGTCGA	150	
Db	62	CGCGCTTGTTCGGCGCATGTGACGGCAGCGGCTCTGC-----CACGCTGAGTGGCTGC	115	
QY	151	GGACTGGGATTTCAATATGCGTCATTAGAGATGACTTTTCAATTCTCCCCCAAGAA	210	
Db	116	GGACTGGGCTTTCGATATGCTTCGGGTTTAGAGCGGATTTCTTCAATTCTCTCCGAAA	175	
QY	211	AAACTGTTTCGGTTTCGGTGAACCTGTGACAGAACTTCCTCAAGTACAAAAAGGTGAA	270	
Db	176	AAACTGTTCGGTTTCGGGAACTGTGACAGAACTTCTACTGAAATACAAAAGGTTGAA	235	
QY	271	CAAAATCACTTTGAGTTTGAAGAACACAGCTGTTAGATCCAGACATAAAGATGACAGA	330	
Db	236	CAAAATGACTTAGAGTTGTGAAGAACACAGCTGCTGATCCTGTATATAAAGGATGACAGA	295	
QY	331	TCATCAACTGGCTGTAGAAATTCGGTCTTATCATGTACTTTCACAAAAAGACTTTTGAGC	390	
Db	296	TCATTAACTGGCTACTAGAAATTCGGTTCCTCTGTCTGATCTTGCACAAAGACTTTTGAGC	355	
QY	391	AACTTATCAGTATTATATTAAGATTGCGTTGGTTGGAATGAAGTCAACAGTAGTGGAA	450	
Db	356	AACTTATAAACATCATATTTGAGATTGCGCTGGTTTGAATAGAAGTCAGAGAGTGGTGGAG	415	
QY	451	AGTATTTGGCTTTTCTGGTAAUCTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTG	510	
Db	416	AGTATTTGGCTTTTCTGGTAAUCTGTGTGTGCACAGACTGTCTTCTTAGACCAATGTC	475	
QY	511	TCAGCATGATTGCTTCCCATTTTGTGCTCCCGAGTGATCATTAAGGAAGCGCATGTAG	570	
Db	476	TCAGCATGATTGCGTGTCTATTTTGTACTCCCGGAGTAATTGTCAAGGAAGTGGCATAG	535	
QY	571	ATGTTTTCAGATTCTGATGATGAAGATGATAATCTTCTCGAAATTTTGACATGTTCACA	630	
Db	536	AUGTTTCAGATTCTGATGACGAAGATGACAAATCTTCTGCAATTTTGTACATGTTCACA	595	
QY	631	GAGCTTTGCAATAATAGCAAGATATGTACATGACACCGTGGTTTCTCATGCCCAATAC	690	
Db	596	GAGCTTTGCAATAATAACAGATATGTCCCATCGACACCATGGTTTCTTAATGCCAATAC	655	



QY 1770 AAGAGGTCAAGAAATTCATGATCTCTATTTATCAGGTATGGGAACATGAGTGTCTGAA 1829  
| | | | |  
Db 1735 AAGAGGTCAAGAAATTCATGATCTCTATTTATCAGGTATGGGAACATGAGTGTCTGAA 1794  
| | | | |  
QY 1830 GAGCTCAGAGTTCAGAAACCCATGAAAGAA-----CATAGTGAAGATGAAGATGA 1885  
| | | | |  
Db 1795 GAGCTTCAGAGTTCAGAAATCTACTTAAAGAAATCTCTCTGCGAAGAGGCAATGG 1854  
| | | | |  
QY 1886 TGACTTCTCAGAGGCGAAGTGCCTCCAGAAATGATACCGTGAT 1927  
| | | | |  
Db 1855 CACATCTCAGAGGGAATCTGCTCCAGAAAGATCTCAAT 1896  
| | | | |

## RESULT 3

Y16973  
LOCUS RNY16973 Rat liver ESTs (E.Olivier) Rattus norvegicus cDNA clone  
DEFINITION RNY16973 Rat liver ESTs (E.Olivier) Rattus norvegicus cDNA clone  
ACCESSION Y16973  
VERSION Y16973.2 GI:4727035  
KEYWORDS EST.  
SOURCE Rattus norvegicus (Norway rat)

## ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 1800)

## REFERENCE

Olivier, E., Soury, E., Risler, J.L., Smith, F., Schneider, K.,  
Lochner, K., Jouzeau, J.Y., Pey, G. and Salier, J.P.  
A novel set of hepatic mRNAs preferentially expressed during an  
acute inflammation in rat represents mostly intracellular proteins  
Genomics 57 (3), 352-364 (1999)  
99263497  
10329001

## COMMENT

On Apr 2, 1998 this sequence version replaced gi:3006069.  
Contact: E. Olivier  
U78 INSERM  
543 chemin de la Breteque, 76233 Bois-Guillaume, France  
POLY-NO.

## FEATURES

## source

Location/Qualifiers  
1..1800  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="IRU573"  
/tissue\_type="liver"  
/dev\_stage="adult"  
/clone\_lib="Rat liver ESTs (E.Olivier)"  
/note="Organ: liver; Vector: lambda ZAPII; Library  
construction: Ripberger, J.A. et al. J. Biol. Chem. (1995)  
270(50):2998-3006"

## ORIGIN

Query Match 39.4%; Score 815.4; DB 14; Length 1800;  
Best Local Similarity 75.7%; Pred. No. 8.2e-223;  
Matches 141; Conservative 0; Mismatches 366; Indels 86; Gaps 29;

QY 135 AAGCTGGCGGTCCAGACTGGGATTTCAATATCGCTGATAGAGAAATGACATTTTC 194  
| | | | |  
Db 23 AGCTGAGTGGCTCGCGACTGGGCTTTCAACATGCTTGCATTAGAGAGTGAATTTCTCA 82  
| | | | |  
QY 195 AATTCTCCCCAAGAAAAACTGTCGGTTGTTGGAACTGTGCAGAGAGTCTTGTCTGAAG 254  
| | | | |  
Db 83 GTCTCCCAAGAAAC-----TGTCGGTGGTGGACTGTGCAGAGCTTACTGAGT 133  
| | | | |  
QY 255 TACAAAAGGGTGAACAAATGACTTTGAGTTGTTGAAAGAACAGCTGTTAGATCCAGAC 314  
| | | | |  
Db 134 ---ACAAAAGGTGAACAAATGACTTAGAGTTGTTGAAGAA-CAGCTGTCTGATCCTGAT 189  
| | | | |  
QY 315 ATAAAGGATGACAGATCATCAACTGGCTGCTAGAAATTCGGTTCTTATCATGCTACTG 374  
| | | | |  
Db 190 ATAGAGGATGATCAGATCATTAACCTGGCTGGAAATTCGTTCTCTCTGCTGACTGTG 249  
| | | | |

QY 375 ACAAAGACTTTGAGCAACTTATCAGTATATATTAAGATTCGCTTGGTTGAATAGAGT 434  
| | | | |  
Db 250 ACAAAGACTTTGAGCAACTTATAAACATCATATTCAGATTCGCTTGGTTGAATAGAGT 309  
| | | | |  
QY 435 CAAACAGTAGTGAAGAGTATTTGGCTTTTCTTGGTAATCTCTATCAGCAGACAGACTGTT 494  
| | | | |  
Db 310 CAGACGGTGGTGGAGAGTATTTGGCTTTTCTTGGTAATCTTGTGTGAGCAGACAGACTGTC 369  
| | | | |  
QY 495 TTCTCAGACCGGTGCTCAGCATGATTCCTCCATTTTGTGCTCCCTCCCGAGTGCATTT 554  
| | | | |  
Db 370 TTCTTAGACCAATGCTCAGCATGATTCCTTCATTTTGTACTCTCCCGAGTATTTGTC 429  
| | | | |  
QY 555 AAGGAAGGGGATGATGATGTTTCAGATTCCTGATGATGAAGATGATATATCTTCCTGCAAT 614  
| | | | |  
Db 430 AAGGAAGGTGGCATTTGATG-TTCAGAGTCTGAGGATGAAGATGACATCTTCCTGCAAT 488  
| | | | |  
QY 615 TTTGACACATGTCACAGAGCGCTTGCATAATATAGCAGATATGTACCATCGACACCGTGG 674  
| | | | |  
Db 489 TTTGACACATGTCACAAAGCGCTTGCAAAATATACAAAGATATGTCCCTTCACACCATGG 548  
| | | | |  
QY 675 TTTCTCATGCCAAATACCTGTTGGAAAAATTTCCATTTTTCGAAAAATCAGAGAAACAATG 734  
| | | | |  
Db 549 TTTCTAATGCCAATACCTGTTAGAAAAATTTCCATTTTGTGAGGAATCAGAGAAACAATG 608  
| | | | |  
QY 735 GAATGTTACGTTTCATAACTTACTAAGATTTAGTGTATATTTTCCAACTTTGAGGCAATGAA 794  
| | | | |  
Db 609 GAATGCTATGTTTCATAACCTATTAAAGGATAAGTTTATATTTCCCAAGTTTGAGCGGTGAA 668  
| | | | |  
QY 795 ATTCTGAGCTTATTTTGAAGAACTACTCAAGTTGATGTGAATGATCCCGCAGGTT 854  
| | | | |  
Db 669 ATTCTCAGCTTATTTTGAAGTTC-ACTCAAGTTAGATGTGAGTGTCCGCGCAGAT 727  
| | | | |  
QY 855 ATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGTGGGACAGATTTCCACGGAAGGA 914  
| | | | |  
Db 728 ATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGTGGGACAGATTTCCACGGAAGGA 787  
| | | | |  
QY 915 TTGTTTAAATATGAGATGAAGATGAAGAACTGAAATGAACAAAGGCTGTCTTGAACGG 974  
| | | | |  
Db 788 CTG-TTAAATATGAGTGA- -GATGAGGAACTGACCTGAGAAAGAGCTGAGCAGAGCGC 844  
| | | | |  
QY 975 CTCGACGAGTGTGATCCCTGATGAGCGCGCGCTGGACATCTCTGATGTCTTTGTTTGG 1034  
| | | | |  
Db 845 CT-GGCGGATGGCTATCCCACTGAGAGGCGCTGGACATCTCTGCTCTCTGCTGCTC 903  
| | | | |  
QY 1035 TCCTACATGAAGGATGTCTGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1094  
| | | | |  
Db 904 TCCTACATGAAGGATGTCTGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 959  
| | | | |  
QY 1095 CTATATCGGACCTGATGAATGAATCTTTGAGAACTCTCTGTTGGCCACCATGCTCTCTGC 1154  
| | | | |  
Db 960 TTATATCGTATCTGATATCTATCTTT---GACAACTGTGTTGGCCACACATGCTCT-CTGC 1015  
| | | | |  
QY 1155 CATGTACAGTCTTTTCATGTTTCTGATGTTTCAAAATTTGGGATTCGACAGAGCAATTT 1214  
| | | | |  
Db 1016 CATGTACAGTCTTTTCATGTTTCTGATGTTTCTGATGTTTCTGATGTTTCTGATGTTTCTGAT 1072  
| | | | |  
QY 1215 TTGAAACATCTCTGGAAGAAATTTGAGGACCCCAAGTATCTCTGATCATCATCAGGAGCT 1274  
| | | | |  
Db 1073 TTGAAACATCTCTGGAAGTGCAGG---TCATAATACCTGTCTCATCATCAG--CAGCT 1126  
| | | | |  
QY 1275 GCTGAAATATATTTGGAAGCTTTTGGCAAGAGCTAAATTTTATCTCTTCTTATCTGTA 1334  
| | | | |  
Db 1127 GCTGAAATATATG---AGCTTTTGGCAGAACTTAA--TTAATCTCTTTATCATCTGTG 1180  
| | | | |  
QY 1335 AAATCATGCTGATGCTTTTGGTTAACTGGCTGGACATATACCTTAAATAACAGGATTCG 1394  
| | | | |  
Db 1181 AAGTCATGCTGGATTC-TTGGGTAACCTGGCTGGAC-TGTACCTTAAATAGCAGGTTGG 1238  
| | | | |  
QY 1395 GGAAACAAAGGATCTGCGATGTTGCTCTCCATGGACCAATTTTATCTAGCCTGCGCAAGCT 1454  
| | | | |  
Db 1239 ACA-----AGGCTTTGTGATGTGCACTCATGGACCAATTTATC---AGCTGCCAAGCT 1287  
| | | | |

```
QY 1455 GTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTGAGCGAAACCTGAAAGAGGT 1514
    |||
Db 1288 GTGT-----CTACACTGTGTTTATGACACAGCACATT-----GAGTGCACCTGAGAGT 1336

QY 1515 TTGCAGTATCTTTCAGAGTCTGAATTTTGGCGGATAGTAGAGCCAGCTTAAATCCCTTG 1574
    |||
Db 1337 CTACGGTATCTCAGAGTCTAAATTTTGAACGCAATGTTGTCAGCCAGCTGAACCCACTG 1396

QY 1575 AAGATTTGCTGCGCTCAGTGGTTAACTTTTGTGCTCAATCAAAATAGTACCAAGTTC 1634
    |||
Db 1397 AAGATCTGCTGCCAGTGTGTAATCTCTGCTGCATACAAATATACAGTGTGTCGCTA 1456

QY 1635 GTCTTGTGTACACCATCATTTGAGAGAAACAATCGCCAGATGCTGCCAGTCAATAGAGT 1694
    |||
Db 1457 CACATCATGAGA-----GGACAATCGCCAGATGCTCCCTGTTAT--CGAGT 1501

QY 1695 ACCGCTGGAGGAGCTCAGTGCAGATCTGCACAAACCCGCTGACACCTTCTTCCCTTT 1754
    |||
Db 1502 ACAGCTGGTGGAGACTCTGTGCAGACCTGCACCAACCCATTTGGACACTTCTTTCCTTT 1561

QY 1755 GATCCCTGTGTGCTGAAGAGGTCAAAAGAAATTCATTGATCCTATTTATCAGGTATGGAA 1814
    |||
Db 1562 GACCCCTGTATGCTTAAGAGGTCAAGAGATTCATCGATCCTGTCTATCAGGCTTGGAA 1621

QY 1815 GACATGAGTGTGAAGAGCTACAGGAGTTCAGAAACCCATCAAAAGACATAGTGGAA 1874
    |||
Db 1622 GACATGAGTGTGAAGAGCTGAGAGCTGAGGAGCT--TAGGAACCCACTAGAAAGCAGCTAGTGGAG 1680

QY 1875 GATGAAGATGATGACTTCTGAAGCGGAGTGGCCAGAAATGATACCGTGTATGGGATC 1934
    |||
Db 1681 GATGAAGACGATGACTTTTGAAGCGGAGTG---CCGAGGTGACACAGTGGCTGGCCCT 1737

QY 1935 ACACCAAGCTCCTTTGACACGCAATTTCCGAAGTTCCTTCAAGTAGTGTGGGCTCCCAACC 1994
    |||
Db 1738 ACCCGCACTCTTTGACACCCACTTCG-AGTCTTCAGTAGTGTGGGCTCCCTCCT 1796

QY 1995 GTG 1997
    |||
Db 1797 GTG 1799
```

```
RESULT 4
BQ214727
LOCUS BQ214727
DEFINITION AGENCOURT_7595062 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5062849
5', mRNA sequence.
ACCESSION BQ214727
VERSION BQ214727.1 GI:20396127
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.G.E. Consortium (LLNL)
Cloning Distribution: MGC clone distribution information can be
found through the I.M.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM1334 row: p column: 18
High quality sequence stop: 595.
Location/Qualifiers
1..896
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
```

## FEATURES

source

```
/clone="IMAGE:6062849"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 2 kb. Library constructed by Life
Technologies."
```

## ORIGIN

```
Query Match 34.4%; Score 711; DB 13; Length 896;
Best Local Similarity 96.5%; Pred. No. 6.8e-193;
Matches 748; Conservative 0; Mismatches 25; Indels 2; Gaps 2;

QY 1001 CGAGCGCTGGACATCTCTGATGTTTGTGTTTGTCTACATGAAGAGTCTGCTATGT 1060
    |||
Db 16 CGCGTCCGTGGACATCTCTGATGTTTGTGTTTGTCTACATGAAGAGTCTGCTATGT 75

QY 1061 AGATGCTAGCTGTGATACGGCAAAACAAGATCTATATCGGACCTGATAAATCTT 1120
    |||
Db 76 AGATGCTAGCTGTGATACGGCAAAACAAGATCTATATCGGACCTGATAAATCTT 135

QY 1121 TGACAAACTCTGTTGCCACCCATGCTCTCCATGTCAGATTTTTCATGTTTACCT 1180
    |||
Db 136 TGACAAACTCTGTTGCCACCCATGCTCTCCATGTCAGATTTTTCATGTTTACCT 195

QY 1181 CTGTAGTTTCAATTTGGGATTCGAGAGCATTTTGGAAACATCTCTGGAAAAAATTGCA 1240
    |||
Db 196 CTGTAGTTTCAATTTGGGATTCGAGAGCATTTTGGAAACATCTCTGGAAAAAATTGCA 255

QY 1241 GGACCAAGTATCTGTCATCATCAGCAGCTCTGGAAATTTATTTGGAAGCTTTT 1300
    |||
Db 256 GGACCAAGTATCTGTCATCATCAGCAGCTCTGGAAATTTATTTGGAAGCTTTT 315

QY 1301 GSCAAGAGCTAAATTTATTTCTTATTACTGTAAATCATGCTAGATCTTTTGGTTAA 1360
    |||
Db 316 GSCAAGAGCTAAATTTATTTCTTATTACTGTAAATCATGCTAGATCTTTTGGTTAA 375

QY 1361 CTGGCTGCACATATACCTTAATACAGGATTCGGAAACAAAGGATTTGCGATTTGC 1420
    |||
Db 376 CTGGCTGCACATATACCTTAATACAGGATTCGGAAACAAAGGATTTGCGATTTGC 435

QY 1421 TCTCATCGACCATTTTACTCAGCTGCCAGCTGTGTTCTTACACCTTTGTTTATAGACA 1480
    |||
Db 436 TCTCATCGACCATTTTACTCAGCTGCCAGCTGTGTTCTTACACCTTTGTTTATAGACA 495

QY 1481 CAAGCAGCTTTTGAGCGGAAACCTGAAAGAGTTTGCAGTATCTTCAGAGTCTGAAATTT 1540
    |||
Db 496 CAAGCAGCTTTTGAGCGGAAACCTGAAAGAGTTTGCAGTATCTTCAGAGTCTGAAATTT 555

QY 1541 TGAGCGGATAGTAGAGCCAGCTAAATCCCTCGAAGATTTGCTGCCCTCAGTGGTTAA 1600
    |||
Db 556 TGAGCGGATAGTAGAGCCAGCTAAATCCCTCGAAGATTTGCTGCCCTCAGTGGTTAA 615

QY 1601 CTTTCTTGTGCTCAATCACAATAAGTACCACTGTCTTCTGCTACA--CCATCATTGAGA 1659
    |||
Db 616 CTTTCTTGTGCTCAATCACAATAAGTACCACTGTCTTCTGCTACA--CCATCATTGAGA 675

QY 1660 GGAACATTCGCGCAGATGCTGCCAGTCAATAGAGTACCGCT--GGAGGAGACTCAGTGCAG 1718
    |||
Db 676 GGAACATTCGCGCAGTATGCTGCCGCTCAATTAAGAGTACCGCTGGGAGGAGACTCAGTGCAG 735

QY 1719 ATCTGCAAAACCCGCTGGACACCTTCTCCCTTTGATCCCTCTGTGCTGAAGA 1773
    |||
Db 736 ATCTGCAAAACCCGCTGGACACCTTCTCCCTTTGATCCCTCTGTGCTGAAGA 790
```

```
RESULT 5
CD359064
LOCUS CD359064
DEFINITION AGENCOURT_14303450 NIH_MGC_180 Homo sapiens cDNA clone
IMAGE:30389999 5', mRNA sequence.
ACCESSION CD359064
```

```

VERSION CD359064.1 GI:31130475
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 894)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM464 row: f column: 24
High quality sequence stop: 506.
FEATURES
    source
    1..894
        Location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:30389999"
            /lab_host="DH10B-ron A (T1 and T5 phage resistances)"
            /clone_lib="NIH_MGC_180"
            /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
            Site 2: EcoRV (destroyed); Library is oligo-dT primed and
            directionally cloned (EcoRV site is destroyed upon
            cloning). Average insert size 1.68 kb. Library was
            constructed by (Invitrogen). Note: this is a NIH_MGC
            Library."
ORIGIN
    Query Match 32.7%; Score 675.6; DB 14; Length 894;
    Best Local Similarity 98.5%; Pred. No. 1.1e-182;
    Matches 703; Conservative 0; Mismatches 9; Indels 2; Gaps 2;
Qy 1152 TGCATGTACAGTGTTCATGTTTACCTCTGTAGTTTCAAAATGGGATTCGCAGAGCA 1211
Db 1 TGCCATGTACAGTGTTCATGTTTACCTCTGTAGTTTCAAAATGGGATTCGCAGAGCA 60
Qy 1212 TTTTGGAACTCTCTGGAAAAAATTCGAGGCCAACCAAGTAATCCGCCATCATCAGGCAG 1271
Db 61 TTTTGGAACTCTCTGGAAAAAATTCGAGGCCAACCAAGTAATCCGCCATCATCAGGCAG 120
Qy 1272 GCTGCTGGAATATATATGGAAAGCTTTTGGCAAGAGCTAAATTTATTCCTTATTACT 1331
Db 121 GCTGCTGGAATATATATGGAAAGCTTTTGGCAAGAGCTAAATTTATTCCTTATTACT 180
Qy 1332 GTAAATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACAGGAT 1391
Db 181 GTAAATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACAGGAT 240
Qy 1392 TCGGGAAACAAGGCATCTCGGATGTTGCTTCGATGGACCAATTTTACTCGCCCTGCCAA 1451
Db 241 TCGGGAAACAAGGCATCTCGGATGTTGCTTCGATGGACCAATTTTACTCGCCCTGCCAA 300
Qy 1452 GCTGTGTTCTACACCTCTGTTTTTATAGACACAGAGCTTTTGAGCGGAAACCTGAAGAA 1511
Db 301 GCTGTGTTCTACACCTCTGTTTTTATAGACACAGAGCTTTTGAGCGGAAACCTGAAGAA 360
Qy 1512 GCTTTGCAGTATCTTCAGAGTCTGAATTTTGACGGGATAGTATGATGAGCCAGCTAATCCC 1571
Db 361 GCTTTGCAGTATCTTCAGAGTCTGAATTTTGACGGGATAGTATGATGAGCCAGCTAATCCC 420
Qy 1572 CTGAAGATTGCTGCCCTCAGTGGTTAACTTTTTTGTGCAATCACAAATAAGTACCAG 1631
Db 421 CTGAAGATTGCTGCCCTCAGTGGTTAACTTTTTTGTGCAATCACAAATAAGTACCAG 480
Qy 1632 CTCGCTCTTCGTACACCATCATTGAGAGGAACAATGCCAGATGCTGCCAGTCAATTAGG 1691
Db 481 CTCGCTCTTCGTACACCATCATTGAGAGGAACAATGCCAGATGCTGCCAGTCAATTAGG 540
Qy 1692 AGTACCCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTTCTCCC 1751
Db 541 AGTACCCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTTCTCCC 600
Qy 1752 TTTGAT-CCCTGTGCTGAGAGCTCAAGAAATTCATTGATCCTATTATCAGGTATG 1810
Db 601 TTTGATCCCTGTGCTGAGAGCTCAAGAAATTCATTGATCCTATTATCAGGTATG 660
Qy 1811 GGAAGACATGAGTCTGCTGAAGAGCTACAGGAGTT-CAAGAAACCCATGAAAAAGG 1863
Db 661 GGAAGACATGAGTCTGGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAG 714
RESULT 6
BG502871 788 bp mRNA linear EST 27-MAR-2001
LOCUS 602550430F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4658071 5',
DEFINITION mRNA sequence.
ACCESSION BG502871
VERSION BG502871.1 GI:13464388
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 788)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1451 row: 1 column: 08
High quality sequence stop: 699.
FEATURES
    source
    1..788
        Location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:4658071"
            /tissue_type="embryonal carcinoma"
            /lab_host="DH10B (T1 phage-resistant)"
            /clone_lib="NIH_MGC_61"
            /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1:
            SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc);
            Double-stranded cDNA was prepared from cell line RNA.
            and 3' adaptors were used in cloning as follows: 5'
            adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
            sequence: 5'-ATTCTAGAGCGCGCCGACATG-dT(30)BN-3'
            (where B = A, C, or G and N = A, C, G, or T). Average
            insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
            contained inserts by PCR. This library was enriched for
            full-length clones and was constructed by Clontech
            Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
            Library."
ORIGIN
    Query Match 32.7%; Score 675.4; DB 12; Length 788;
    Best Local Similarity 96.1%; Pred. No. 1.1e-182;
    Matches 756; Conservative 0; Mismatches 23; Indels 8; Gaps 6;
```



1807	Qy	TATGGGAAGCATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAGGACA	1866
428	Db	TGTGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAGGACA	487
1867	Qy	TAGTGGAAAGATGAAGATGATGACATTTCTGAAAGGCGAAGTGCCGCCAGAAATGATACCGTGA	1926
488	Db	TAGTGGAAAGATGAAGATGATGACATTTCTGAAAGGCGAAGTGCCGCCAGATGATACCGTGA	547
1927	Qy	TTGGGATCACACAAAGTCCTTTTGACACCGCATTTTCGGAAGTCTCTTCAAGTAGTGTTGGGCT	1986
548	Db	TTGGGATCACACAAAGTCCTTTTGACACCGCATTTTCGGAAGTCTCTTCAAGTAGTGTTGGGCT	607
1987	Qy	CCCCACCCGTTGTGTACATGCAACCCCAAGTCCCTCTCAGCGCGCAAAATTTGTGTACTGAGA	2046
608	Db	CCCCACCCGTTGTGTACATGCAACCCCAAGTCCCTCTCAGCGCGCAAAATTTGTGTACTGAGA	667
2047	Qy	TGTGACATTTGGGAT	2061
668	Db	TGTGACATTTGGGAT	682

RESULT 8	
BI561523	
LOCUS	733 bp mRNA linear EST 05-SEP-2001
DEFINITION	603256222F1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:5298581 5', mRNA sequence.
ACCESSION	BI561523
VERSION	BI561523.1 GI:15448837
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 733)
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-re@mail.nih.gov">cgapbs-re@mail.nih.gov</a> Tissue Procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshituki and Piero Carninci (RIKEN) CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing Array: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLAM11755 row: 1 column: 06 High quality sequence stop: 717.

```

FEATURES
source
1. .733
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5298581"
/lab_host="DH10B"
/clone_lib="NIH_MGC 97"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript Ks+); Site_1: BamHI; Site_2: SalI-XhoI
(Gcgag); Oligo-dt primed using primer
5'-TTTTTTTTTTTTTNN-3', size-selected for average
insert size 2.2 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIMH/NHGRI, National Institutes of Health). Note: this is
a NIH_MGC Library."

ORIGIN
Query Match 32.2%; Score 666; DB 12; Length 733;
Best Local Similarity 97.5%; Pred. No. 5.6e-180;
Matches 697; Conservative 0; Mismatches 16; Indels 2; Gaps 2;

```

Qy	53	GC	TTAGTCTCGCCGCAATGCGCGCACCGCTCGTTCACACGCGTTTCCGGGAGATGCGGC	112
Db	7	GGG	TTAGTCTCGCCGCAATGCGCGCACCGCTCGTTCACACGCGTTTCCGGGAGATGCGGC	66
Qy	113	CGCTT	CGTCTCTCGCAGTTAAGAAAGCTGGGCGGTCGAGGACTGGGATTTCAAATATGCG	172
Db	67	CGCTT	CGTCTCTCGCAGTTAAGAAAGCTGGGCGGTCGAGGACTGGGATTTCAAATATGCG	126
Qy	173	TGC	ATTAGAAATGACATTTTTCAAATCTCCCCAGAAAACTGTTTCGGTTTGGTGGAAAC	232
Db	127	TGC	ATTAGAAATGACATTTTTCAAATCTCCCCAGAAAACTGTTTCGGTTTGGTGGAAAC	186
Qy	233	TGT	GACAGAAGTCTTGTGAGTACAAAAAGGGTGAAACAATGACTTTTGAGTCTGTGAA	292
Db	187	TGT	GACAGAAGTCTTGTGAGTACAAAAAGGGTGAAACAATGACTTTTGAGTCTGTGAA	246
Qy	293	GA	ACCAGCTGTTAGATCCAGACATAAAGGATGACAGATCATCACTCGCTGCTAGAATT	352
Db	247	GA	ACCAGCTGTTAGATCCAGACATAAAGGATGACAGATCATCACTCGCTGCTAGAATT	306
Qy	353	CGG	TCTTCTCATGTACTTTGACAAAAGACTTTTGAGCAACTTATCAGTATTATATTAA	412
Db	307	CGG	TCTTCTCATGTACTTTGACAAAAGACTTTTGAGCAACTTATCAGTATTATATTAA	366
Qy	413	ATT	GCCTTGGTGTGAATAGAAGTCAAAACAGTAGTGGAGAGATATTGGCTTTTCTTGGTAA	472
Db	367	ATT	GCCTTGGTGTGAATAGAAGTCAAAACAGTAGTGGAGAGATATTGGCTTTTCTTGGTAA	426
Qy	473	TC	TGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATGCTTCCCATTT	532
Db	427	TC	TGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATGCTTCCCATTT	486
Qy	533	TGT	GCCTCCCGAGTGTATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCCTGATGCA	592
Db	487	TGT	GCCTTCCCGAGTGTATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCCTGATGCA	546
Qy	593	AG	ATGATAATCTTCCCTGAAATTTTGACACATGTCCAGAGCCTTGCAAAATATAGCAAG	652
Db	547	AG	ATGATAATCTTCCCTGCAAA-TTTGAACACATGTCCAGAGCCTTGCAAAATATAGCAAG	605
Qy	653	AT	ATGTACCATCGACACCGT--GGTTTCTCATGCCAATACTGCTGCAAAAATTTCCATTG	711
Db	606	AT	ATGTACCATCGACACCGTGGGTTTCTCATGCCAATACTNGTGGAAAAATTTCCATT	665
Qy	712	TT	GAAAAATCAGAGAACACTGGAATGTTACGTTTCATACTTACTAAGGATTAG	766
Db	666	GTT	CGAAATCAGAGAACACTGGAATGTTACGTTTCATACTTACTAAGGATTAG	720

```

RESULT 9
BF203684
LOCUS
DEFINITION
BF203684
996 bp mRNA linear EST 06-NOV-2000
601866331F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:409398 5',
mRNA sequence.
ACCESSION
BF203684
VERSION
BF203684.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 996)
N1 (bases 1 to 996)
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

```

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCM967 row: n column: 07  
High quality sequence stop: 684.

## FEATURES

RESULT 10  
BE734354  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BE734354 757 bp mRNA linear EST 15-SEP-2000  
601565596F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3840623 5',  
mRNA sequence.  
BE734354  
BE734354.1 GI:10148346  
EST.  
Homo sapiens (human)  
Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC

## ORIGIN

Query Match 31.6%; Score 653.6; DB 10; Length 996;  
Best Local Similarity 97.0%; Pred. No. 2.5e-176;  
Matches 687; Conservative 0; Mismatches 19; Indels 2; Gaps 2;

QY 1202 CGCAGAGGCATTTTGGAACTCTCTGGAAAAATTCAGGACCCCAAGTAATCCTGCCAT 1261  
DB 2 CGCAGAGGCATTTTGGAACTCTCTGGAAAAA-TGCAGGACCCCAAGTAATCCTGCCAT 60  
QY 1262 CATCAGGAGGCTGCTGGAAATTAATTCGAAGCTTTTGGCAAGAGCTAAATTAATTC 1321  
DB 61 CATCAGGAGGCTGCTGGAAATTAATTCGAAGCTTTTGGCAAGAGCTAAATTAATTC 120  
QY 1322 TCTTTACTGTAAATCATGCTAGATCTTTTGGTTACTTGCTGCTGCACATACCTTAA 1381  
DB 121 TCTTTACTGTAAATCATGCTAGATCTTTGTTTAACTGGCTGCACATACCTTAA 180  
QY 1382 TAACAGGATTCGGGAACAAAGCATTCTCGCATGTTGCTCCATGGACCAATTTATC 1441  
DB 181 TAACAGGATTCGGGAACAAAGCATTCTCGCATGTTGCTCCATGGACCAATTTATC 240  
QY 1442 AGCCTGCCAGCTGTGTTCTACACCTTTGTTTGAACAACAAGAGCTTTTGAAGGAAA 1501  
DB 241 AGCCTGCCAGCTGTGTTCTACACCTTTGTTTGAACAACAAGAGCTTTTGAAGGAAA 300  
QY 1502 CCTGAAAGAGGTTTGCATATCTTCAGAGTCTGAATTTTGAGCGATAGTATGACCA 1561  
DB 301 CTTGAAAGAGGTTTGCATATCTTCAGAGTCTGAATTTTGAGCGATAGTATGACCA 360  
QY 1562 GCTAAATCCCTGAAGATTGCTGCTCAGTGGTTAACTTTTGTGCTGCAATCACAA 1621  
DB 361 GCTAAATCCCTGAAGATTGCTGCTCAGTGGTTAACTTTTGTGCTGCAATCACAA 420  
QY 1622 TAAGTACAGCTGCTGCTGCTACACATCATTTGAGAGGAACAATGCGCAGATGCTGCC 1681  
DB 421 TAAGTACAGCTGCTGCTGCTACACATCATTTGAGAGGAACAATGCGCAGATGCTGCC 480  
QY 1682 AGTCATTTAGAGTACCGCTGGAGGAGACTCAGTGCAGATCTCGCAAAACCCGCTGACAC 1741  
DB 481 AGTCATTTAGAGTACCGCTGGAGGAGACTCAGTGCAGATCTCGCAAAACCCGCTGACAC 540  
QY 1742 CTTCTCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAGAAATTCATTGATCCTATTTA 1801  
DB 541 CTTCTCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAGAAATTCATTGATCCTATTTA 600  
QY 1802 TCAGTATGGGAGACATGCTGCTGAGAGCTACAGGATTCAGAAACCCATGAATA 1861  
DB 601 TCAGTATGGGAGACATGCTGCTGAGAGCTACAGGATTCAGAAACCCATGAATA 660  
QY 1862 GGACATAGTGAAGATGAAGATGATGACTTCTGAAAGCGGAAGTGCC 1909  
DB 661 GGACATAGTGAAGATGAAGATGATG-CAATCTGAAGAGGATGGCC 707

## FEATURES

1. .757  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3840623"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 21"  
/note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match 31.2%; Score 646; DB 10; Length 757;  
Best Local Similarity 95.9%; Pred. No. 3.3e-174;  
Matches 727; Conservative 0; Mismatches 25; Indels 6; Gaps 6;  
QY 607 CTGCAAAATTTGACACATGTCACAGAGCCTTGCAAATAATACAGAGATATGTACCATCGA 666  
DB 1 CTGCAAAATTTGACACATGTCACAGAGCCTTGCAAATAATACAGAGATATGTACCATCGA 60  
QY 667 CACCGTG-CTTTCTCATGCCAATACTGTTGGAAAAATTTCCATTTGTTGAAAAATCAGAG 725  
DB 61 CACCGTGATTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTGAAAAATCAGAG 120  
QY 726 AGAACATCTGGA-ATGTTACGTTTCACTTACTTAAGAGATAGTATATTTTCCAACTT 784  
DB 121 AGAACATCTGGAATGTTACGTTTCACTTACTTAAGAGATAGTATATTTTCCAACTT 180  
QY 785 GAGGCATGAAATCTCGAGAGCTTATTATTGAAAACTCTCAAGTTGGATGTGAATGCATC 844  
DB 181 GAGGCATGAAATCTCGAGAGCTTATTATTGAAAACTCTCAAGTTGGATGTGAATGCATC 240  
QY 845 CCGCAGAGG-ATTGGAAGATGCTGAAGAAAACAGCAACTCAAACTTGTGTGGGACAGATT 903  
DB 241 CCGCAGAGGCTATTGAAGATGCTGAAGAAAACAGCAACTCAAACTTGTGTGGGACAGATT 300  
QY 904 CCACGAGAGGATTGTTTAATATGAGATGAAGAACTGAACATGAAACAAA-GGCT 962

Db 301 CCACGAGGATGTTTAAATGATGAAGATGAAGAAATGACATGAACAAACGGCT 360  
QY 963 GGTCTCTGAACGGCTCGACAGATGGTGCATCTCTGTAGCCGAGCGCTGGACATCCTGATG 1022  
Db 361 GGTCTCTGAACGGCTCGACAGATGGTGCATCTCTGTAGCCGAGCGCTGGACATCCTGATG 420  
QY 1023 TCTTTGGTTTGGTCCCTACATGAGAGATGCTGCTGATGATGATGATGATGATGATGATG 1082  
Db 421 TCTTTGGTTTGGTCCCTACATGAGAGATGCTGCTGATGATGATGATGATGATGATGATGATG 480  
QY 1083 AAAACAAAGATCTATATCGCGACTGATAAACAATCTTTGACAAATCCTGTGCCCCACC 1142  
Db 481 AAAACAAAGATCTATATCGCGACTGATAAACAATCTTTGACAAATCCTGTAGCCCAAC 540  
QY 1143 CATGCCCTCCTGCCATGTACAGTATTTTTCATGTTTACCTCTGTAGTTTCAAATGGGATTC 1202  
Db 541 CATGCCCTCCTGCCATGTACAGTATTTTTCATGTTTACCTCTGTAGTTTCAAATGGGATTC 600  
QY 1203 GCAGAGCATTTTGGAACTCTCTGGAATAATTTGAGACCCCAAGTAATCTTGCCATC 1262  
Db 601 GCAGAGCATTTTGGAACTCTCTGGAATAATTTGAGACCCCAAGTAATCTTGCCATC 660  
QY 1263 ATCAGGCGAGCTGCTGAAATTAATTTGGAAGCTTTTGGCAAGACTAAATTTATTCCT 1322  
Db 661 ATCAGGCGAGCTGCTGAAATTAATTTGGAAGCTTTTGGCAAGACTAAATTTATTCCT 719  
QY 1323 CTTATTACTGTAAATCATGCTAGATCTTTTGGTTAA 1360  
Db 720 C-TATTACTGTCCATCATGCTAGAAATCTTAGGATAA 756

RESULT 11  
BG428305 1037 bp mRNA linear EST 14-MAR-2001  
LOCUS 602498983f1 NIH\_MGC\_75 Homo sapiens cDNA clone IMAGE:4612605 5',  
DEFINITION mRNA sequence.

ACCESSION BG428305  
VERSION BG428305.1 GI:13334811  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1037)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L10M361 row: e column: 22  
High quality sequence stop: 637.  
Location/Qualifiers

FEATURES  
source 1..1037  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4612605"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH\_MGC\_75"  
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site: 1:  
SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc); 5' and  
3' adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCGCGCGGCAGATG-dt(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.65  
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones  
and was constructed by Clontech Laboratories (Palo Alto,  
CA). Note: this is a NIH\_MGC Library."

Query Match 30.8%; Score 637.4; DB 12; Length 1037;  
Best Local Similarity 88.7%; Pred. No. 1.2e-171;  
Matches 797; Conservative 0; Mismatches 81; Indels 21; Gaps 9;

QY 49 GGTCTCTGAACGGCTCGACAGATGGTGCATCTCTGTAGCCGAGCGCTGGACATCCTGATG 108  
Db 2 GGTCTCTGAACGGCTCGACAGATGGTGCATCTCTGTAGCCGAGCGCTGGACATCCTGATG 61  
QY 109 CGCGCGCTTCTGCTCTGCAAGTAAAGAGCTGGCGCTCGAGACTGGGATTCAAATA 168  
Db 62 CGCGCGCTTCTGCTCTGCAAGTAAAGAGCTGGCGCTCGAGACTGGGATTCAAATA 121  
QY 169 TGGTGCATTTAGGAATGACTTTTCAATTTCTCCCAAGAAAGAACTGTTGGTTTGGTG 228  
Db 122 TGGTGCATTTAGGAATGACTTTTCAATTTCTCCCAAGAAAGAACTGTTGGTTTGGTG 181  
QY 229 GAACTGTGACAGAGTCTTCTGCAAGTAAAGAGCTGAAACAAATGCACTTTGAGTTGT 288  
Db 182 GAACTGTGACAGAGTCTTCTGCAAGTAAAGAGCTGAAACAAATGCACTTTGAGTTGT 241  
QY 289 TGAAGAACCGAGCTGTTAGATCCAGACATAAAGAGTACCAAGATCATCAACTGGCTGCTAG 348  
Db 242 TGAAGAACCGAGCTGTTAGATCCAGACATAAAGAGTACCAAGATCATCAACTGGCTGCTAG 301  
QY 349 AATTCGGTCTCTTATCATGCTTACGAAAGAACTTTGAGCAACTTATCAGTATTATAT 408  
Db 302 AATTCGGTCTCTTATCATGCTTACGAAAGAACTTTGAGCAACTTATCAGTATTATAT 361  
QY 409 TAGAATTCCTTGGTTCGAATGAAAGTCAAAACAGTAGTGAAGAGTATTGGCTTTCTTG 468  
Db 362 TAGAATTCCTTGGTTCGAATGAAAGTCAAAACAGTAGTGAAGAGTATTGGCTTTCTTG 421  
QY 469 GTAATCTTGTATCAGCACAGACTGTTTCTCAGACCGTGTCTCAGCATGATTCCTCCC 528  
Db 422 GTAATCTTGTATCAGCACAGACTGTTTCTCAGACCGTGTCTCAGCATGATTCCTCCC 481  
QY 529 ATTCTTGTGCTCCCG-AGTGATCATTAAGAGAGCGGATGTAGATTTTCAGATTCGTAT 587  
Db 482 ATTCTTGTGCTCCCGAGAGTGATCAATTAAGAGAGCGGATGTAGATTTTCAGATTCGTAT 541  
QY 588 GATGAAGATGATTAATCTTCTGCAAAATTTGACACATGTCCAGAGCGCTTGCATAATAATA 647  
Db 542 GATGAAGATGATTAATCTTCTGCAAA-TTTGACACATGTCCAGAGCGCTTGCATAATAATA 600  
QY 648 GCAAGATGATGATCAATCGACACCG-TGGTTTCTCATGCCAATATCTGG--TGGAAAAATTT 704  
Db 601 GCAAGATGATGATCAATCGACACCGTTGGTTTCTCATGCCAATATCTGGGTGCGCAAAATTC 660  
QY 705 CCATTTCTTCGAAA---TCAGAGAGAACACTGGAATGTT--ACGTTTCAT--AACTTAC 756  
Db 661 CCATATGCTCAGAAAAAATCTCCAGAGAGAACACTGGAATGTTTCACCGTTTCAGAACTTACT 720  
QY 757 TAAGGATTAAGTATATTTTCCAACTTG--AGGCATGAAATTTCTGGAGCTTATTATGA 814  
Db 721 TCAGGATCAAGGATATATGTCAGAGCTTGGAGGATGCAAAATTTCTGGAGCTTATTATGA 780  
QY 815 AAAACTACTCAAGTTGGATGTGAATCATCCCGGACGGGTATTGAAGATG-----CTGA 868  
Db 781 AAACTACTCAAGTTGGATGTGAATGCTCCCGGGCGGGGTATTTTGAAGACTGCTGAGG 840  
QY 869 AGAAACAGCAACTCAAACTTGTGGTGGGACAGATTTCCACGGAGGATTTTATATGG 927  
Db 841 AAAGCAGAAATACAAACTGTCGGGAAACAAATCCCGGCAAGGGATTTGTCAGTTGG 899

RESULT 12  
BU510623 889 bp mRNA linear EST 12-SEP-2002  
LOCUS BU510623





of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

## ORIGIN

```
Query Match      30.3%; Score 627.2; DB 13; Length 1122;
Best Local Similarity 98.5%; Pred. NO. 1.1e-168;
Matches 644; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

QY 13 GCTGAAGGAGCTGGCGCTCGGCGTGGAGCGAGCGTGGTGGTGGTGGCGCCCAATGG 72
Db 1 GCTGGAAGGAGCTGGCGCTCGGCGTGGAGCGAGCGTGGTGGTGGTGGCGCCCAATGG 60

QY 73 CGGCGCGGCTGTTTCCACGCGCTTTCCGCGGAGATGCGGCGCGCTTCTGCTCTGCGAGTTA 132
Db 61 CGGCGCGGCTGTTTCCACGCGCTTTCCGCGGAGATGCGGCGCGCTTCTGCTCTGCGAGTCA 120

QY 133 AGAAGCTGGCGCTGCGAGGACTGGGATTTCAATATGCGTGCTCATAGAGAATGACTTTT 192
Db 121 AGATGCTGGCGCTGCGAGGACTGGGATTTCAATATGCGTGCTCATAGAGAATGATTTTT 180

QY 193 TCAATTTCTCCCAAGAAAACCTGTTTCGTTTGGTGGAACTGTGACAGAGTCTTGTCTGA 252
Db 181 TCAATTTCTCCCAAGAAAACCTGTTTCGTTTGGTGGAACTGTGACAGAGTCTTGTCTGA 240

QY 253 AGTCAAAAAGGTTGAAACAAATGACTTTGAGTTTCTTGAAGAACCGAGCTGTTAGATCCAG 312
Db 241 AGTCAAAAAGGTTGAAACAAATGACTTTGAGTTTCTTGAAGAACCGAGCTGTTAGATCCAG 300

QY 313 ACATPAAGGATGACAGATCATCACTGCTGCTAGAAATTCGTTCTTCTATCATGTACT 372
Db 301 ACATPAAGGATGACAGATCATCACTGCTGCTAGAAATTCGTTCTTCTATCATGTACT 360

QY 373 TGACAAAAGACTTTGAGCAACTTATCAGTATTATTAAGATTGCTTGGTTGATGATGAA 432
Db 361 TGACAAAAGACTTTGAGCAACTTATCAGTATTATTAAGATTGCTTGGTTGATGATGAA 420

QY 433 GTCAAAAGTGTGGAAGAGTATTTGGCTTTCTTGGTAAATCTTGTATCAGCACAGACTG 492
Db 421 GTCAAAAGTGTGGAAGAGTATTTGGCTTTCTTGGTAAATCTTGTATCAGCACAGACTG 480

QY 493 TTTTCTCAGACCGGTCTCAGCATGATGCTTCCCATTTTGTGCTCCCGAGTGATCA 552
Db 481 TTTTCTCAGACCGGTCTCAGCATGATGCTTCCCATTTTGTGCTCCCGAGTGATCA 540

QY 553 TTAAGGAAGGCGATGTAGATGTTTTCAGATTTCTGATGATGAAGATGATATCTTCTGCAA 612
Db 541 TTAAGGAAGGCGATGTAGATGTTTTCAGATTTCTGATGATGAAGATGATATCTTCTGCAA 600

QY 613 ATTTTGACAGTGCACAGAGCGCTTGCAAAATTAAGCAAGATATGTACCATCGA 666
Db 601 ATTTTGACAC--GTCACAGAGCGCTTGCAAAATTAAGCAAGATATGTACCATCGA 652
```

RESULT 14  
AV703279 671 bp mRNA linear EST 09-OCT-2000  
LOCUS AV703279 ADB Homo sapiens cDNA clone ADBCCD12 5', mRNA sequence.  
ACCESSION AV703279  
VERSION AV703279.1 GI:10720608  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 671)  
Peng, Y., Song, H., Huang, Q., Huang, C., Gu, Y., Yang, Y., Gao, G.,  
Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z.,  
Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S.,  
Zhong, M., Lu, G., Hu, R., Chen, J., Chen, Z. and Han, Z.  
Homo sapiens cDNA ADB clones  
Unpublished (2000)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

## FEATURES

## source

```
1..671
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ADBCCD12"
/tissue_type="Adrenal gland"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="ADB"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
```

## ORIGIN

```
Query Match      29.9%; Score 617.8; DB 9; Length 671;
Best Local Similarity 97.0%; Pred. NO. 4.1e-166;
Matches 650; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

QY 416 GCCTTGGTTGAATAGAGTCAAAACAGTAGTAGTGAAGAGTATTGGCTTTTCTTGGTAAATCT 475
Db 1 GCCTTGGTTGAATAGAGTCAAAACAGTAGTAGTGAAGAGTATTGGCTTTTCTTGGTAAATCT 60

QY 476 TGATATCAGCACAGACTGTTTTCCTCAGACCGGTGCTCAGCATGATTGCTTCCCATTTTGT 535
Db 61 TGATATCAGCACAGACTGTTTTCCTCAGACCGGTGCTCAGCATGATTGCTTCCCATTTTGT 120

QY 536 GCCTTCCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAGATTCGTATGATGAAGA 595
Db 121 GCCTTCCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAGATTCGTATGATGAAGA 180

QY 596 TGATAATCTTCTCTCAAAATTTTGACATGTGCACAGAGCCCTGCAAAATAGCAGAGATA 655
Db 181 TGATAATCTTCTCTCAAAATTTTGACATGTGCACAGAGCCCTGCAAAATAGCAGAGATA 240

QY 656 TGTAACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGAAATAATTTCCATTTTGTTCG 715
Db 241 TGTAACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGAAATAATTTCCATTTTGTTCG 300

QY 716 AAAATCAGAGAGAACACTGGAAATGTTAGTTTCATAAATTACTAAGGATTAGTATATTT 775
Db 301 AAAATCAGAGAGAACACTGGAAATGTTAGTTTCATAAATTACTAAGGATTAGTATATTT 360

QY 776 TCCAACTTGGAGGCATGAAATTTCTGGAGCTTATTATTGAAAACTACTCAAGTTGGATGT 835
Db 361 TCCAACTTGGAGGCATGAAATTTCTGGAGCTTATTATTGAAAACTACTCAAGTTGGATGT 420

QY 836 GAATGCATCCCGCAGGCTATTTGAAGATGCTGAAAGAAACAGCAACTCAAACTTGTGGTGG 895
Db 421 GAATGCATCCCGCAGGCTATTTGAAGATGCTGAAAGAAACAGCAACTCAAACTTGTGGTGG 480

QY 896 GACAGATTCACGGAAGGATTTGTTTATATGATGGAAGATGAAGAACTGAAACATGAACA 955
Db 481 GACAGATTCACGGAAGGATCGTTTAAATATGATGGAAGATGAAAGATGAAACATCAAC 540

QY 956 AAAGGCTGG--TCCGTGAACCGCTGACCAAGATGGTGCATCTGTAGCCGAGCGCTTGACA 1014
Db 541 ACAGGCTGNTTCCGTGAACCGCTGACCAAGATGGTGCATCTGTAGCCGAGCGCTTGACA 600

QY 1015 TCCTGATGCTCTTTGGTTTGTTCCTACATGAAGGATGTCTGTATGTAGATGTTAAGGTG 1074
Db 601 TCCTGATGCTCTTTGGTTTGTTCCTACATGAAGGATGTCTGTATGTAGATGTTAAGGTG 659

QY 1075 ATAAGGCAAA 1084
Db 660 ATACCGCCCA 669
```

```
RESULT 15
BI464721
LOCUS      603202341F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5268011 5',
DEFINITION mRNA sequence.
ACCESSION BI464721
VERSION   BI464721.1 GI:15255364
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: rgabs-r@mail.nih.gov
           Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
           cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
           Toshiyuki and Piero Carninci (RIKEN)
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LNL at:
           http://image.lnl.gov
           Plate: LLAM11676 row: b column: 12
           High quality sequence stop: 725.
FEATURES
         Location/Qualifiers
             1..736
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /clone="IMAGE:5268011"
                 /lab_host="DH10B"
                 /clone_lib="NIH MGC 97"
                 /note="Organ: testis; Vector: pBluescriptR (modified
                 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
                 (gtcag); Oligo-dT primed using primer
                 5'-TTTTTTTTTTTTVN-3', size-selected for average
                 insert size 2.2 kb and normalized to ROT 5. This is a
                 primary library enriched for full-length clones and
                 constructed using the Cap-trapper method (Carninci, in
                 preparation). Library constructed by M. Brownstein
                 (NIH/NHGRI, National Institutes of Health). Note: this is
                 a NIH_MGC Library."
ORIGIN
Query Match      29.7%; Score 614.6; DB 12; Length 736;
Best Local Similarity 95.4%; Pred. No. 3.6e-165;
Matches 699; Conservative 0; Mismatches 24; Indels 10; Gaps 6;

QY      34  GGCTCAGGCGCAGCGGTGCGTGTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGC 93
Db      6  GGCTTCAGGCGCAGCGGTGCGGTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGC 65

QY      94  GTTTGCGGGAGATGCGGGCGCTTCGTCTCTCGAGTTAAGAGCTGGGCGGCTCGAGGA 153
Db      66  GTTTGCGGGAGATGCGGGCGCTTCGTCTCTCGAGTTAAGAGCTGGGCGGCTCGAGGA 125

QY      154 CTGGGATTTCAATATGCGTGCAATTAGAGATGACATTTTCAATTCTCCCCCAAGAAAAA 213
Db      126 CTGGGATTTCAATATGCGTGCAATTAGAGATGACATTTTCAATTCTCCCCCAAGAAAAA 185

QY      214 CTGTTCCGTTTGGTGAAGCTGTGACAGAACTTTGCTGAAGTACA--AAAAGGTTGAAC 271
Db      186 CTGTTCCGTTTGGTGAAGCTGTGACAGAACTTTGCTGAAGTACAACACAGGTTGAAC 245

QY      272 AAATGACTTTGAGTTGTTGAAGAACCGAGCTGTTAGATCCAGACATAAAGATGACCAGAT 331
Db      246 AAATGACTTTGAGTTGTTGAAGAACCGAGCTGTTAGATCCAGACATAAAGATGACCAGAT 305
```

```
QY      332 CATCACTGGCTGTAGAAATCCGTTCTTATCATGTACTTGACAAAAAGACTTTTGAGCA 391
Db      306 CATCACTGGCTGTAGAAATCCGTTCTTATCATGTACTTGACAAAAAGACTTTTGAGCA 365

QY      392 ACTTATCAGTATTATATTAAAGATTGCCCTTGGTTGAATAGAACTCAACACAGTAGTAGGAAGA 451
Db      366 ACTTATCAGTATTATATTAAAGATTGCCCTTGGTTGAATAGAACTCAACACAGTAGTAGGAAGA 425

QY      452 GTATTGGCTTTTCTTTGGTAAATCTTGATCAGCACAGACTGTTTCCCTCAGACCGTGCT 511
Db      426 GTATTGGCTTTTCTTTGGTAAATCTTGATCAGCACAGACTGTTTCCCTCAGACCGTGCT 485

QY      512 CAGCATGATTGCTTCCCATTTTGTGCTCCCG-AGTGATCATTAAGGAAGCGGATGT-A 569
Db      486 CAGCATGATTGCTTCCCATTTTGTGCTCCCGAAGTGATCATTAAGGAAGCGGATGTAA 545

QY      570 GATGTTTCAGATTCTGATGATGAAGATGATAATCTTCTGCAAAATTTTGACACA----TG 625
Db      546 GATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCGTGCAAAATTTTGACACAAGTG 605

QY      626 TCACAGAGCCTTGCAAAATTAATAGCAAGATATGTACCATCGACACCGTGTTTCTCATGCC 685
Db      606 TCACAGAGCCTTG-AAAATAATAGCAAGATATGTACCATCGACA-CGTGGTTTCTCATGCC 663

QY      686 AATACTGGTGGAAAAATTTCCATTTGTTGAAAAATCAGAGAGAACACTGGAATGTTACGT 745
Db      664 AATACTGGTGGAAAAATTTCCATTTGTTGAAAAATCAGAGAGAACACTGGAATGTTACGT 723

QY      746 TCATAACTTTACTA 758
Db      724 TCATAAATTACTA 736
```

Search completed: July 18, 2004, 06:23:00  
Job time : 5413 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 04:52:59 ; Search time 931 Seconds  
(without alignments)  
10836.727 Million cell updates/sec

Title: US-09-932-678-1  
Perfect score: 2068  
Sequence: 1 acagagctgtggtggaag.....tgacattgggattcccat 2068

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 319092 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/2/pubnpa/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubnpa/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubnpa/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubnpa/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubnpa/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubnpa/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubnpa/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubnpa/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubnpa/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubnpa/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubnpa/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubnpa/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubnpa/US09\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubnpa/US10A\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubnpa/US10B\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubnpa/US10C\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubnpa/US10\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubnpa/US60\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubnpa/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2068	100.0	2068	9	US-09-932-678-1
2	2018.4	97.6	2040	9	US-09-932-013A-1
3	1457.6	70.5	1770	16	US-10-094-466-7
4	567	27.4	2493	16	US-10-094-749-544
5	420.8	20.3	1461	13	US-10-221-625-172
6	230	11.1	2103	9	US-09-925-301-192
7	192	9.3	535	15	US-10-029-386-10613
8	188.8	9.1	535	15	US-10-029-386-12002
9	184	8.9	184	15	US-10-029-386-24313
10	182.4	8.8	184	15	US-10-029-386-25702
11	156	7.5	557	9	US-09-998-598-1177
12	153.2	7.4	598	15	US-10-029-386-7910
13	146.6	7.1	225	9	US-09-783-590-11086
14	140.6	6.8	506	15	US-10-029-386-8224

15	139.6	6.8	342	15	US-10-029-386-21924
16	116.8	5.6	140	9	US-09-728-445-420
17	106.6	5.2	414	9	US-09-864-761-4848
18	93.4	4.5	95	9	US-09-864-761-21583
c	84.2	4.1	89	15	US-10-029-386-21610
20	61.6	3.0	634	13	US-10-027-632-263885
21	61.6	3.0	634	13	US-10-027-632-263886
22	61.6	3.0	634	13	US-10-027-632-263887
23	61.6	3.0	634	13	US-10-027-632-263888
24	61.6	3.0	634	16	US-10-027-632-263885
25	61.6	3.0	634	16	US-10-027-632-263886
26	61.6	3.0	634	16	US-10-027-632-263887
27	61.6	3.0	634	16	US-10-027-632-263888
28	61.2	3.0	634	13	US-10-027-632-263889
29	61.2	3.0	634	16	US-10-027-632-263889
30	49.4	2.4	51	13	US-10-393-815-96
31	49.4	2.4	51	13	US-10-393-815-97
32	48.2	2.3	12770	16	US-10-397-569-4
33	47.6	2.3	1671	15	US-10-032-585-6770
34	39.2	1.9	593	13	US-10-424-599-80209
35	39.2	1.9	1819	13	US-10-424-599-135210
36	39.2	1.9	3108	15	US-10-193-764-68
37	39.2	1.9	4937	14	US-10-092-880-3
38	39.2	1.9	4937	15	US-10-193-764-66
39	39.2	1.9	684707	16	US-10-398-221-9
c	39.2	1.9	3011208	16	US-10-398-221-2058
41	38.6	1.9	638	9	US-09-974-300-6328
42	38.6	1.9	3284	13	US-10-424-599-90356
c	37.8	1.8	2418	15	US-10-349-680-153
44	37.8	1.8	2910	15	US-10-193-764-31
45	37.8	1.8	2928	15	US-10-193-764-29

ALIGNMENTS

RESULT 1  
US-09-932-678-1  
; Sequence 1, Application US/09932678  
; Patent No. US2002090706A1  
; GENERAL INFORMATION:  
; APPLICANT: Reeder, Ronald H.  
; APPLICANT: Moorefield, Beth  
; APPLICANT: Greene, Elizabeth A.  
; TITLE OF INVENTION: HUMAN RN3 AND COMPOSITIONS AND METHODS RELATING  
; TITLE OF INVENTION: THERETO  
; FILE REFERENCE: 14538A-005810US  
; CURRENT APPLICATION NUMBER: US/09/932.678  
; PRIOR FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: 60/225,893  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2068  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-932-678-1

Query Match	100.0%;	Score 2068;	DB 9;	Length 2068;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2068;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ACAGAGCTGTGGCTGGAAGAGCTGGGCACTCCGCCCTGAGGCGCAGCGGTGCGTTAGT	60	
Db	1	ACAGAGCTGTGGCTGGAAGAGCTGGGCACTCCGCCCTGAGGCGCAGCGGTGCGTTAGT	60	
QY	61	TCGCGCCCAATGGCGGACCGCTGCTTACACACGCGTTTGGCGGAGATGCGCGCTTCGT	120	
Db	61	TCGCGCCCAATGGCGGACCGCTGCTTACACACGCGTTTGGCGGAGATGCGCGCTTCGT	120	
QY	121	CCTCTGCAGTTAAGAGCTGGGCGCTCGAGGACTGGGATTTCAAAATATCGTGCTTAG	180	

Db 121 CCTCTGCGATTAAAGAGCTGGCGCGCTCGAGGACTGGGATTTCAAATATGCGTGCATTAG 180  
Qy 181 AGAATGACTTTTTCAATTTCTCCCCCAAGAAAAAAGTTCGGTTTGGTGGAACTGTGACAG 240  
Db 181 AGAATGACTTTTTCAATTTCTCCCCCAAGAAAAAAGTTCGGTTTGGTGGAACTGTGACAG 240  
Qy 241 AAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAGC 300  
Db 241 AAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAGC 300  
Qy 301 TGTTAGATCCAGACATAAAGGATGACCATCAATCGGCTGTAGAAATTCGGTTCTT 360  
Db 301 TGTTAGATCCAGACATAAAGGATGACCATCAATCGGCTGTAGAAATTCGGTTCTT 360  
Qy 361 CTATCATGTACTTGTACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCCTT 420  
Db 361 CTATCATGTACTTGTACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCCTT 420  
Qy 421 GGTTCGAATAGAGTCAAAAGAGTGGAGAGTATTGGCTTTCTTGSTAACTCTGTAT 480  
Db 421 GGTTCGAATAGAGTCAAAAGAGTGGAGAGTATTGGCTTTCTTGSTAACTCTGTAT 480  
Qy 481 CAGCAGACAGTCTTTTCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCCTC 540  
Db 481 CAGCAGACAGTCTTTTCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCCTC 540  
Qy 541 CCCGAGTGATCATTAAGGAAGCGATGTAGATTGTTTTCAGATCTGATGATGAAGATGATA 600  
Db 541 CCCGAGTGATCATTAAGGAAGCGATGTAGATTGTTTTCAGATCTGATGATGAAGATGATA 600  
Qy 601 ATCTTCTCGCAATTTTGACATGTCACAGAGCCTTGCAATATAGCAAGATATGTAC 660  
Db 601 ATCTTCTCGCAATTTTGACATGTCACAGAGCCTTGCAATATAGCAAGATATGTAC 660  
Qy 661 CATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTTCGAAAT 720  
Db 661 CATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTTCGAAAT 720  
Qy 721 CAGAGAGAACATCTGAATGTAGTTTCAATACTTAAGGATTAAGTATATTTTCCAA 780  
Db 721 CAGAGAGAACATCTGAATGTAGTTTCAATACTTAAGGATTAAGTATATTTTCCAA 780  
Qy 781 CCTTCAGGATGAATTTCTGAGCTTATTATTGAAAACTACTCAAGTTGGATGTCAATG 840  
Db 781 CCTTCAGGATGAATTTCTGAGCTTATTATTGAAAACTACTCAAGTTGGATGTCAATG 840  
Qy 841 CATCCCGCAGGATATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGTGGACAG 900  
Db 841 CATCCCGCAGGATATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGTGGACAG 900  
Qy 901 ATTCCAAGGATTTGTTTATATGATGAAGATGAAGAACTGAACATGAAGAAACAAAGG 960  
Db 901 ATTCCAAGGATTTGTTTATATGATGAAGATGAAGAACTGAACATGAAGAAACAAAGG 960  
Qy 961 CTGCTCTGAACCGCTCGACAGATGTCATCTGTAGCCGAGCGCTGGACATCTCTGA 1020  
Db 961 CTGCTCTGAACCGCTCGACAGATGTCATCTGTAGCCGAGCGCTGGACATCTCTGA 1020  
Qy 1021 TGCTTTTGGTTTGTCTCATAGAGGATGTCTGCTATGTAGTGGTAAAGTTGATAACG 1080  
Db 1021 TGCTTTTGGTTTGTCTCATAGAGGATGTCTGCTATGTAGTGGTAAAGTTGATAACG 1080  
Qy 1081 GCAAAACAAAGGATCTATATCGGACCTGATAAACAATCTTTTGACAAACTCCTGTGCCCA 1140  
Db 1081 GCAAAACAAAGGATCTATATCGGACCTGATAAACAATCTTTTGACAAACTCCTGTGCCCA 1140  
Qy 1141 CCCATGCTCCTGCCATGTACAGTTTTCATGTTTTTACCTCTGTAGTTTCAAAATGGGAT 1200  
Db 1141 CCCATGCTCCTGCCATGTACAGTTTTCATGTTTTTACCTCTGTAGTTTCAAAATGGGAT 1200  
Qy 1201 TCGCAGAGGCAATTTTGGAAACATCTCTCGAAAAAAATTCGAGGCCCAAGTAATCTCGCCA 1260  
Db 1201 TCGCAGAGGCAATTTTGGAAACATCTCTCGAAAAAAATTCGAGGCCCAAGTAATCTCGCCA 1260

Qy 1261 TCATCAGCAGCGCTCTGGAATATATATTTGAAAGCTTTTGGCAAGACTAAATTTATTC 1320  
Db 1261 TCATCAGCAGCGCTCTGGAATATATATTTGAAAGCTTTTGGCAAGACTAAATTTATTC 1320  
Qy 1321 CTCCTTATTACTGTAATAATCATGCCCTAGATCTTTTGGTTAACTGGCTGACATATACCTTA 1380  
Db 1321 CTCCTTATTACTGTAATAATCATGCCCTAGATCTTTTGGTTAACTGGCTGACATATACCTTA 1380  
Qy 1381 ATAAACAGGATTCGGGAAACAAAGGCAATTTCTGCGATTTCTCCATGGACCATTTTACT 1440  
Db 1381 ATAAACAGGATTCGGGAAACAAAGGCAATTTCTGCGATTTCTCCATGGACCATTTTACT 1440  
Qy 1441 CAGCCCTCGAAGCTGTGTTCTACACTTTGTTTTTAGACACAGCAGCTTTTGGCGGAA 1500  
Db 1441 CAGCCCTCGAAGCTGTGTTCTACACTTTGTTTTTAGACACAGCAGCTTTTGGCGGAA 1500  
Qy 1501 ACCTCAAAAGAGGTTTGCAGTATCTTCAGAGTCTGAAATTTTGGCGGATGATGAGCC 1560  
Db 1501 ACCTCAAAAGAGGTTTGCAGTATCTTCAGAGTCTGAAATTTTGGCGGATGATGAGCC 1560  
Qy 1561 AGCTAAATCCCTGAAGATTTGCCCTGAGTCTTCTGCTACCAATCATTTGAGAGGAACTTTCGCTGCAATCACA 1620  
Db 1561 AGCTAAATCCCTGAAGATTTGCCCTGAGTCTTCTGCTACCAATCATTTGAGAGGAACTTTCGCTGCAATCACA 1620  
Qy 1621 ATAAAGTACCAAGCTCGTCTTCTGCTACCAATCATTTGAGAGGAACTTTCGCTGCAATCACA 1680  
Db 1621 ATAAAGTACCAAGCTCGTCTTCTGCTACCAATCATTTGAGAGGAACTTTCGCTGCAATCACA 1680  
Qy 1681 CAGTCATTAGGAGTACCGCTGGAGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACA 1740  
Db 1681 CAGTCATTAGGAGTACCGCTGGAGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACA 1740  
Qy 1741 CTTCTTCCCTTTGATCCCTGTGCTGAAGAGTCAAGAAATTCATTGATCCTATT 1800  
Db 1741 CTTCTTCCCTTTGATCCCTGTGCTGAAGAGTCAAGAAATTCATTGATCCTATT 1800  
Qy 1801 ATCAGGTATGGAAGACATGAGTGTGTAAGAGCTACAGAGTTCAGAAACCCCATCAAAA 1860  
Db 1801 ATCAGGTATGGAAGACATGAGTGTGTAAGAGCTACAGAGTTCAGAAACCCCATCAAAA 1860  
Qy 1861 AGGACATAGTGAAGATGAAGATGATGACTTTCTGAAGCGGAAAGTGCCTCCAGATGATA 1920  
Db 1861 AGGACATAGTGAAGATGAAGATGATGACTTTCTGAAGCGGAAAGTGCCTCCAGATGATA 1920  
Qy 1921 CCGTGAATTGGATCACCACCAAGCTCTTTGACACGCAATTCGAAAGTCTTCAAGTAGTG 1980  
Db 1921 CCGTGAATTGGATCACCACCAAGCTCTTTGACACGCAATTCGAAAGTCTTCAAGTAGTG 1980  
Qy 1981 TGCGCTCCCAACCGTGTGTATCATGCAACCCAGTCCCTCTGACGGCAGAAATTTGTGA 2040  
Db 1981 TGCGCTCCCAACCGTGTGTATCATGCAACCCAGTCCCTCTGACGGCAGAAATTTGTGA 2040  
Qy 2041 CTGAGATGTGACATTTGGGATTCCTCAT 2068  
Db 2041 CTGAGATGTGACATTTGGGATTCCTCAT 2068

RESULT 2  
US-09-952-013A-1  
; Sequence 1, Application US/09952013A  
; Patent No. US20020146801A1  
; GENERAL INFORMATION:  
; APPLICANT: GRUMMET, Ingrid  
; APPLICANT: VINGRON, Martin  
; TITLE OF INVENTION: RNA POLYMERASE I TRANSCRIPTION FACTOR TIP-1A  
; FILE REFERENCE: 38485-0007  
; CURRENT APPLICATION NUMBER: US/09/952, 013A  
; CURRENT FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: PCT/DE00/00767  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: DE 199 11 992.9  
; PRIOR FILING DATE: 1999-03-17

```
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2040
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (40)..(1992)
; OTHER INFORMATION:
; NAME/KEY: misc feature
; LOCATION: (721)..(721)
; OTHER INFORMATION: n is a, c, g or t
; NAME/KEY: misc feature
; LOCATION: (1435)..(1435)
; OTHER INFORMATION: n is a, c, g or t
US-09-952-013A-1

Query Match          97.6%; Score 2018.4; DB 9; Length 2040;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2019; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 47 GCGGTCCGTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 106
Db 18 GCGGTCCGTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 77

Qy 107 TGGCGCCGCTTCGTCCTCTGCAGTTAAGAAAGCTGGCGCGCTCGAGGACTGGGATTTCAA 166
Db 78 TGGCGCCGCTTCGTCCTCTGCAGTTAAGAAAGCTGGCGCGCTCGAGGACTGGGATTTCAA 137

Qy 167 TATCGGTGCATATAGAAATGACTTTTCAATCTCCCGCAAGAAACTGTTGCGGTTGG 226
Db 138 TATCGGTGCATATAGAAATGACTTTTCAATCTCCCGCAAGAAACTGTTGCGGTTGG 197

Qy 227 TGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAGSGGTGAACAAATGACTTTGAGTT 286
Db 198 TGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAGSGGTGAACAAATGACTTTGAGTT 257

Qy 287 GTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCACTGGCTGCT 346
Db 258 GTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCACTGGCTGCT 317

Qy 347 AGAATTCGGTCTTCTATCATGTACTTGACAAAAGACTTTGACAACTTATCAGTATTAT 406
Db 318 AGAATTCGGTCTTCTATCATGTACTTGACAAAAGACTTTGACAACTTATCAGTATTAT 377

Qy 407 ATTAAGATTGCTTGGTTGAATAGAAGTCAAAACAGTAGTGAAGATTTTGGCTTTTCT 466
Db 378 ATTAAGATTGCTTGGTTGAATAGAAGTCAAAACAGTAGTGAAGATTTTGGCTTTTCT 437

Qy 467 TGGTAAATCTTGATCAGCACAGACTGTTTTCTCAGACCGTGTCTCAGCATGATTGCTTC 526
Db 438 TGGTAAATCTTGATCAGCACAGACTGTTTTCTCAGACCGTGTCTCAGCATGATTGCTTC 497

Qy 527 CCATTTTGTGCTCCCGAGTATCATTTAAGGAAGGAGATAGATGTTTTCAGATTCTGA 586
Db 498 CCATTTTGTGCTCCCGAGTATCATTTAAGGAAGGAGATAGATGTTTTCAGATTCTGA 557

Qy 587 TGATGAAGATGATTAATCTTCCTGCAAAATTTTGAACATGTCCAGAGAGCTTTGCAAAAT 646
Db 558 TGATGAAGATGATTAATCTTCCTGCAAAATTTTGAACATGTCCAGAGAGCTTTGCAAAAT 617

Qy 647 AGCAAGATATGATCATCGACACCGTGGTTTCTCATGCCAATACTGTTGGAATAATTTC 706
Db 618 AGCAAGATATGATCATCGACACCGTGGTTTCTCATGCCAATACTGTTGGAATAATTTC 677

Qy 707 ATTTGTTGAAAATCAGAGAAACACTGGAATGTTACGTTTATTAATCTTACTAAGATTAG 766
Db 678 ATTTGTTGAAAATCAGAGAAACACTGGAATGTTTACGTTTATTAATCTTACTAAGATTAG 737

Qy 767 TGTATATTTTCCAACTTGGAGCATGAATTTCTGGAGCTTATTTGAAAACACTACTCAA 826
Db 738 TGTATATTTTCCAACTTGGAGCATGAATTTCTGGAGCTTATTTGAAAACACTACTCAA 797
```

```
Qy 827 GTTGGATGTGAATGCATCCCGCAGGAGTATTGAAGATGCTGAAGAAACAGCAACTCAAC 886
Db 798 GTTGGATGTGAATGCATCCCGCAGGAGTATTGAAGATGCTGAAGAAACAGCAACTCAAC 857

Qy 887 TTGTGGTGGACAGATTCACGGAAGGATTTTAATATATGATGAAGATGAAGAACTGA 946
Db 858 TTGTGGTGGACAGATTCACGGAAGGATTTTAATATATGATGAAGATGAAGAACTGA 917

Qy 947 ACATGAACAAAGGCTGGTCTCGAACGGCTCGACAGATGGTGCATCTCTGAGCCGAGCG 1006
Db 918 ACATGAACAAAGGCTGGTCTCGAACGGCTCGACAGATGGTGCATCTCTGAGCCGAGCG 977

Qy 1007 CCTGGACATCCTGATGCTTTTGGTTTTGCTCTACATGAAGGATGTCTGCTATGTAGATGG 1066
Db 978 CCTGGACATCCTGATGCTTTTGGTTTTGCTCTACATGAAGGATGTCTGCTATGTAGATGG 1037

Qy 1067 TAAGTTGATAACGGCAAAAACAAAGGATCTATATTCGGACCTGATAAACATCTTTGCAA 1126
Db 1038 TAAGTTGATAACGGCAAAAACAAAGGATCTATATTCGGACCTGATAAACATCTTTGCAA 1097

Qy 1127 ACTCCTGTTGCCCAACCCATCGCTCCTGCCATGTACAGTCTTTTATGTTTTTACCTCTGAG 1186
Db 1098 ACTCCTGTTGCCCAACCCATCGCTCCTGCCATGTACAGTCTTTTATGTTTTTACCTCTGAG 1157

Qy 1187 TTTCAAATTTGGGATTCGACAGGCAATTTTGGAAACATCTCTCGGAAAAAATTCGAGGACCC 1246
Db 1158 TTTCAAATTTGGGATTCGACAGGCAATTTTGGAAACATCTCTCGGAAAAAATTCGAGGACCC 1217

Qy 1247 AAGTAATCCTGCCATCATCAGGAGGCTGCTGGAAATTTATTTGGAAGCTTTTGGCAAG 1306
Db 1218 AAGTAATCCTGCCATCATCAGGAGGCTGCTGGAAATTTATTTGGAAGCTTTTGGCAAG 1277

Qy 1307 AGCTAAATTTATTTCTCTTACTGTAAATCATGCTAGATCTTTTGGTTAACTGGCT 1366
Db 1278 AGCTAAATTTATTTCTCTTACTGTAAATCATGCTAGATCTTTTGGTTAACTGGCT 1337

Qy 1367 GCATATATCCTTAATAACCCAGGATTCGGAAACAAAGGCAATCTCGGATTTGCTCTCCA 1426
Db 1338 GCATATATCCTTAATAACCCAGGATTCGGAAACAAAGGCAATCTCGGATTTGCTCTCCA 1397

Qy 1427 TGGACCAATTTACTCAGGCTGCCAAGCTGTGCTTACACCTTTGTTTTAGACAACAGCA 1486
Db 1398 TGGACCAATTTACTCAGCTGCCAAGCTGTGCTTACNCCCTTTGTTTTAGACAACAGCA 1457

Qy 1487 GCTTTTTCAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCG 1546
Db 1458 GCTTTTTCAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCG 1517

Qy 1547 GATAGTATGAGCCAGCTAAATCCCTGGAAGATTTGCTGCGCTCAGTGGTTAACTTTTT 1606
Db 1518 GATAGTATGAGCCAGCTAAATCCCTGGAAGATTTGCTGCGCTCAGTGGTTAACTTTTT 1577

Qy 1607 TGCTGCAATCACAAATAAGTACAGCTCGCTTCTGCTTACACCATCATTCAGAGGAAACAA 1666
Db 1578 TGCTGCAATCACAAATAAGTACAGCTCGCTTCTGCTTACACCATCATTCAGAGGAAACAA 1637

Qy 1667 TCGCCAGATGCTGCGAGTCAATTAGAGTACCGCTGGAGGAGACTCAGTCAGATCTGCAC 1726
Db 1638 TCGCCAGATGCTGCGAGTCAATTAGAGTACCGCTGGAGGAGACTCAGTCAGATCTGCAC 1697

Qy 1727 AAACCCGCTGACACCTTTTCCCTTTGATCCCTGTGTGCTGAAGGTTCAAGAAATTT 1786
Db 1698 AAACCCGCTGACACCTTTTCCCTTTGATCCCTGTGTGCTGAAGGTTCAAGAAATTT 1757

Qy 1787 CATTTGATCTTATTTATCAGGTATGGGAAGACATGAGTCTGAAGGTTCAAGAAATTT 1846
Db 1758 CATTTGATCTTATTTATCAGGTATGGGAAGACATGAGTCTGAAGGTTCAAGAAATTT 1817

Qy 1847 GAAACCCATGAAAGGACATAGTGGGAAGATGAAGATGATGACATTTCTGAAAGGCGAAGT 1906
Db 1818 GAAACCCATGAAAGGACATAGTGGGAAGATGAAGATGATGACATTTCTGAAAGGCGAAGT 1877
```

```
QY 1907 GCCCAGATGATACCGTGATTGGGATCACCAAGCTCCTTTGACAGCATTTCCGAAG 1966
Db 1878 GCCCAGATGATACCGTGATTGGGATCACCAAGCTCCTTTGACAGCATTTCCGAAG 1937
QY 1967 TCCCTCAAGTAGTGTGGGCTCCCAACCGCTGTGTGATGCAACCCAGTCCCTCTGAGC 2026
Db 1938 TCCCTCAAGTAGTGTGGGCTCCCAACCGCTGTGTGATGCAACCCAGTCCCTCTGAGC 1997
QY 2027 GCAGAAATTTGACTGAGATGATGACATTTGGGATTTCCCAT 2068
Db 1998 GCAGAAATTTGACTGAGATGATGACATTTGGGATTTCCCAT 2039

RESULT 3
US-10-094-466-7
; Sequence 7, Application US/10094466
; Publication No. US20030203363A1
; GENERAL INFORMATION:
; APPLICANT: Spytex et al.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM
; TITLE OF INVENTION: AND METHODS OF USING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 21402-290D
; CURRENT APPLICATION NUMBER: US/10/094,466
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/288,148
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/338,375
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/275,579
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/335,302
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/275,601
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/276,000
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/277,338
; PRIOR FILING DATE: 2001-03-20
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patin 2.1
; SEQ ID NO 7
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (40)..(1723)
US-10-094-466-7
```

```
Query Match 70.5%; Score 1457.6; DB 16; Length 1770;
Best Local Similarity 86.2%; Pred. No. 0;
Matches 1743; Conservative 0; Mismatches 9; Indels 270; Gaps 1;

QY 47 GCGGTCGGTTAGTTTCGGCCCAATGGCGGCACCGCTGCTTACACGCGTTTGGCCGGAGA 106
Db 18 GCGGTCGGTTAGTTTCGGCCCAATGGCGGCACCGCTGCTTACACGCGTTTGGCCGGAGA 77
QY 107 TGGGCGCCCTTCGCTCTCTGCAAGTTAAGAGCTGGGCGCGTCGAGGACTGGGATTTCAA 166
Db 78 TGGGCGCCCTTCGCTCTCTGCAAGTTAAGAGCTGGGCGCGTCGAGGACTGGGATTTCAA 137
QY 167 TATGGGTGCATTAGAGATGACTTTTCAATTTCTCCCAAGAAAACTGTTGCGTTTGG 226
Db 138 TATGGGTGCATTAGAGATGACTTTTCAATTTCTCCCAAGAAAACTGTTGCGTTTGG 197
```

```
QY 227 TGGAACTGTGACAGAGTCTTCTGAACTACAAAAGGCTGAAACAAATGACTTTGAGTT 286
Db 198 TGGAACTGTGACAGAGTCTTCTGAACTACAAAAGGCTGAAACAAATGACTTTGAGTT 257
QY 287 GTTGAAGAACCAAGCTGTAGATCCAGACATATAAGGATGACCAAGATCATCAACTGGCTGCT 346
Db 258 GTTGAAGAACCAAGCTGTAGATCCAGACATATAAGGATGACCAAGATCATCAACTGGCTGCT 317
QY 347 AGAATTCGGTCTTCTATCANTACTTGTGACAAAAGCTTTGAGCAACTTTATCAGTATTAT 406
Db 318 AGAATTCGGTCTTCTATCANTACTTGTGACAAAAGCTTTGAGCAACTTTATCAGTATTAT 377
QY 407 ATTAAGATTGCTTGGTTGAATAGAGTCAACACACTAGTGGAGAGTATTTCGCTTTTCT 466
Db 378 ATTAAGATTGCTTGGTTGAATAGAGTCAACACACTAGTGGAGAGTATTTCGCTTTTCT 437
QY 467 TGGTAATCTTGTATCAGCACAGACTGTTTCTCTCAGACCGTCTCTCAGCATGATTGCTTTC 526
Db 438 TGGTAATCTTGTATCAGCACAGACTGTTTCTCTCAGACCGTCTCTCAGCATGATTGCTTTC 497
QY 527 CCATTTTGTGCTCCCGAGTGTATTAAGGAGCGGATGATGTTTTCAGTCTGA 586
Db 498 CCATTTTGTGCTCCCGAGTGTATTAAGGAGCGGATGATGTTTTCAGTCTGA 557
QY 587 TGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTACAGAGCCTTCGAATAAT 646
Db 558 TGATGAAGATGATAATCTTCTGCAAAATTTTGACACATGTACAGAGCCTTCGAATAAT 617
QY 647 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCAATPACTGGTGGAAAAATTTCC 706
Db 618 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCAATPACTGGTGGAAAAATTTCC 677
QY 707 ATTTGTTGAAAAATCAGAGAGAACTCTGGAATGTTACGTTTCTAATACTTACTAAGGATTAG 766
Db 678 ATTTGTTGAAAAATCAGAGAGAACTCTGGAATGTTACGTTTCTAATACTTACTAAGGATTAG 737
QY 767 TGTATATTTTCCAACTTTGAGCATGAAATTTCTGAGCTTATTATTGAAAACTACTCAA 826
Db 738 TGTATATTTTCCAACTTTGAGCATGAAATTTCTGAGCTTATTATTGAAAACTACTCAA 797
QY 827 GTTGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAAAACAGCAACTCAAC 886
Db 798 GCTGGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAAAACAGCAACTCAAC 857
QY 887 TTGTGTCGGACAGATTTCCAGGAGGATTTTAAATATGATGATGAAGATGAAGAACTGA 946
Db 858 TTGTGTCGGACAGATTTCCAGGAGGATTTTAAATATGATGATGAAGATGAAGAACTGA 895
QY 947 ACATGAAACAAAAGGCTGGTCTCTGAAACGGCTCGACAGATGGTGCATCCTGTAGCCGAGCG 1006
Db 896 ----- 895
QY 1007 CCTGGACATCCTGATGTTTGGTTTTGTCCTTACATGAAGGATGTCCTCTATGTAGATGG 1066
Db 896 ----- 895
QY 1067 TAAGTTGATAACGGCAAAAACAAAGGATCTATATCGGACCTGATAAACAATCTTGACAA 1126
Db 896 ----- 895
QY 1127 ACTCCTGTTGCCCCCATGCTCCTGCCATGTACAGTTTTTCATGTTTACCTCTGTAG 1186
Db 896 ----- 895
QY 1187 TTTCAAATTTGGGATTCGACAGGCAATTTTGGAAACATCTCTGGAATAATATATGGAAGCTTTTGGCAG 1246
Db 896 -----TGGGATTCGACAGGCAATTTTGGAAACATCTTTGGAATAATATATGGAAGCTTTTGGCAG 947
QY 1247 AAGTAATCCTCCCATCATCAGCAGGCTGCTGGAAAAATATATGGAAGCTTTTGGCAG 1306
Db 948 AAGTAATCCTCCCATCATCAGCAGGCTGCTGGAAAAATATATATGGAAGCTTTTGGCAG 1007
QY 1307 AGCTAAATTTTCTCTTATTACTGTAAATCATGCTAGATCTTTTGGTTAACTGGCT 1366
```



```
Db 1008 AGCTAAATTTATTTCTCTTATTACTGATAAAACCAATGCTAGATCTTTGGTTAACTGGCT 1067
Qy 1367 GCACATATACCTTAAATACCAAGGATTCGGGAACAAAGGATTCCTGCGATGTTGCTCTCCA 1426
Db 1068 GCACATATACCTTAAATACCAAGGATTCGGGAACAAAGGATTCCTGCGATGTTGCTCTCCA 1127
Qy 1427 TGGACCAATTTTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCA 1486
Db 1128 TGGACCAATTTTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCA 1187
Qy 1487 GCTTTTCAGCGGAACCTGGAAGAGGTTTCAGATCTTCAGAGTCTGAAATTTTGAGCG 1546
Db 1188 GCTTTTCAGCGGAACCTGGAAGAGGTTTCAGATCTTCAGAGTCTGAAATTTTGAGCG 1247
Qy 1547 GATAGTGATGAGCCAGCTAAATCCCTGAAGATTTGCCCTGCCCTCAGTGGTTAACTTTTT 1606
Db 1248 GATAGTGATGAGCCAGCTAAATCCCTGAAGATTTGCCCTGCCCTCAGTGGTTAACTTTTT 1307
Qy 1607 TGCTGCAATCAAAATAGTACAGCTGCTGTTCTTCTGCTACACCAATCATTTAGAGGAACAA 1666
Db 1308 TGCTGCAATCAAAATAGTACAGCTGCTGTTCTTCTGCTACACCAATCATTTAGAGGAACAA 1367
Qy 1667 TCGCCAGATGCTGCCAGTCATTAGAGTACCGCTGGAGGAGCTCAGTGCAGATCTGCAC 1726
Db 1368 TCGCCAGATGCTGCCAGTCATTAGAGTACCGCTGGAGGAGCTCAGTGCAGATCTGCAC 1427
Qy 1727 AAACCCGCTGACACCTTTCTTCCCTTTGATCCCTGTGCTGAAGAGGTCAAGAAATTT 1786
Db 1428 AAACCCGCTGACACCTTTCTTCCCTTTGATCCCTGTGCTGAAGAGGTCAAGAAATTT 1487
Qy 1787 CATTGATCCCTATTATACGTTATGGGAAGACATGAGTGTCTGAAGAGGTACAGGATTCAA 1846
Db 1488 CATTGATCCCTATTATACGTTATGGGAAGACATGAGTGTCTGAAGAGGTACAGGATTCAA 1547
Qy 1847 GAAACCCATGAAGAAGACATAGTGGAGATGAAGATGATGATCTTCTGAAAGCGGAGT 1906
Db 1548 GAAACCCATGAAGAAGACATAGTGGAGATGAAGATGATGATCTTCTGAAAGCGGAGT 1607
Qy 1907 GCCCCAGAAATGATACCGTGTGATGGGATCACACCAAGCTCCCTTGACACGCAATTTCCGAAG 1966
Db 1608 GCCCCAGAAATGATACCGTGTGATGGGATCACACCAAGCTCCCTTGACACGCAATTTCCGAAG 1667
Qy 1967 TCCTTCAAGTGTGGCTCCCAACCGTGTGTGATCATGCAACCCAGTCCCTCTGACG 2026
Db 1668 TCCTTCAAGTGTGGCTCCCAACCGTGTGTGATCATGCAACCCAGTCCCTCTGACG 1727
Qy 2027 GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTTCCCCAT 2068
Db 1728 GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTTCCCCAT 1769
```

RESULT 4

US-10-094-749-544

```
; Sequence 544, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: HONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMYUKI
```

```
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 544
; LENGTH: 2493
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-544
```

Query Match

27.4%; Score 567; DB 16; Length 2493;  
Best Local Similarity 98.3%; Pred. No. 3.9e-160;  
Matches 573; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy 1196 GGGATTGCGCAGAGGCAATTTTGGAAACATCTCTGGAATAAATTGCGAGACCCCAAGTAATCC 1255
Db 711 GGGATTGCGCAGAGGCAATTTTGGAAACATCTTTGGAAAAAATTCAGGATCCAGTAATCC 770
Qy 1256 TGGCATCATCAGCGAGGCTGCTGGAATAATTAATTGGAAGCTTTTGGCAAGAGCTAAATT 1315
Db 771 TGGCATCATCAGCGAGGCTGCTGGAATAATTAATTGGAAGCTTTTGGCAAGAGCTAAATT 830
Qy 1316 TATTCTCTCTATTATCTGTAATAATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATA 1375
Db 831 TATTCTCTCTATTATCTGTAATAACCATGCTAGATCTTTTGGTTAACTGGCTGCACATATA 890
Qy 1376 CTTTAAATAACAGGATTCGGGAACAAAGGCAATTCCTGCGATGTTCTCTCCATGGACCAATT 1435
Db 891 CTTTAAATAACAGGATTCGGGAACAAAGGCAATTCCTGCGATGTTCTCTCCATGGACCAATT 950
Qy 1436 TTACTCAGCTCCCAAGCTGTGTTCTACACCTTTGTTTTAGACACAGCAGCTTTTGAG 1495
Db 951 TTACTCAGCTCCCAAGCTGTGTTCTACACCTTTGTTTTAGACACAGCAGCTTTTGAG 1010
Qy 1496 CGGAAACCTGAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGCAGGATAGTAT 1555
Db 1011 CGGAAACCTGAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGCAGGATAGTAT 1070
Qy 1556 GAGCCAGCTAAATCCCTGGAAGATTTGCCCTCAGTGGTTAACTTTTGTGCTGCAAT 1615
Db 1071 GAGCCAGCTAAATCCCTGGAAGATTTGCCCTCAGTGGTTAACTTTTGTGCTGCAAT 1130
Qy 1616 CACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTCAGAGGAACAAATCGCCAGAT 1675
Db 1131 CACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTCAGAGGAACAAATCGCCAGAT 1190
Qy 1676 GCTGCCAGTCAATTAGAGATACCGCTGGAGAGACTCAGTGCAGATCTGCACAAACCCGCT 1735
Db 1191 GCTGCCAGTCAATTAGAGATACCGCTGGAGAGACTCAGTGCAGATCTGCACAAACCCGCT 1250
Qy 1736 GGACACCTTCTCCCTTTGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1778
Db 1251 GGACACCTTCTCCCTTTGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1293
```

RESULT 5

US-10-221-625-172

```
; Sequence 172, Application US/10221625
; Publication No. US20040033942A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: YUE, Henry
; APPLICANT: LAL, Preeti
```





```
QY 1389 GATTCGGGAACAAGGCATTCTGCATGTTGCTCTCCATGACACCATTTTACTCAGCCTGC 1448
Db 61 GATTCGGGAACAAGGCATTCTGCATGTTGCTCTCCATGACACCATTTTACTCAGCCTGC 120
QY 1449 CAAGCTGTGTTCTACACCTTTGTTTATAGACACAAGCAGCTTTTGGAGGGAAACCTGAAA 1508
Db 121 CAAGCTGTGTTCTACACCTTTGTTTATAGACACAAGCAGCTTTTGGAGGGAAACCTGAAA 180
QY 1509 GAAG 1512
Db 181 GAAG 184

RESULT 11
US-09-998-598-1177/c
; Sequence 1177, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meaguer, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1177
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1177

Query Match 7.5%; Score 156; DB 9; Length 557;
Best Local Similarity 100.0%; Pred. No. 6e-36;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1913 GAATGATACCGTGATTGGGATCACACCAAGCTCCCTTTGACAGCATTTCCGAAGTCCTTC 1972
Db 557 GAATGATACCGTGATTGGGATCACACCAAGCTCCCTTTGACAGCATTTCCGAAGTCCTTC 498
QY 1973 AAGTAGTGTGGCTCCCAACCGTGTTGTACATGCAACCCAGTCCCTCTGACGGCAGAA 2032
Db 497 AAGTAGTGTGGCTCCCAACCGTGTTGTACATGCAACCCAGTCCCTCTGACGGCAGAA 438
QY 2033 ATTTGTGACTGAGATGTGACATTTGGGATTTCCCAT 2068
Db 437 ATTTGTGACTGAGATGTGACATTTGGGATTTCCCAT 402

RESULT 12
US-10-029-386-7910/c
; Sequence 7910, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7910
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 17
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 23
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 19
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 33
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 15
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 13
; OTHER INFORMATION: SWISSPROT HIT: Q07436, EVALUE 5.10e-01
; OTHER INFORMATION: NT HIT: g114779150, EVALUE 3.00e-83
; OTHER INFORMATION: EST_HUMAN HIT: BG485267.1, EVALUE 0.00e+00
US-10-029-386-7910

Query Match 7.4%; Score 153.2; DB 15; Length 598;
Best Local Similarity 98.1%; Pred. No. 4.5e-35;
Matches 155; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCCTGAGCGCAGCGGTGCGCTTAGT 60
Db 517 ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCCTGAGCGCAGCGGTGCGCTTAGT 458
QY 51 TCGGCCCAATGCGGCACGCTGCTTTCACACGCTTTCGGGAGATGCGCGCCTTCGT 120
Db 457 TCGGCCCAATGCGGCACGCTGCTTTCACACGCTTTCGGGAGATGCGCGCCTTCGT 398
QY 121 CCTCTGCACTTAAGAAGCTGGGCGCGCTCGAGGACTGGG 158
Db 397 CCTCTGCACTTAAGAAGCTGGGCGCGCTCGAGGACTGGG 360

RESULT 13
US-09-783-590-11086
; Sequence 11086, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11086
; LENGTH: 225
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (45)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (48)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (72)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (75)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (94)
```

```

, OTHER INFORMATION: n equals a,t,g, or c
, NAME/KEY: misc feature
, LOCATION: (153)
, OTHER INFORMATION: n equals a,t,g, or c
, NAME/KEY: misc feature
, LOCATION: (154)
, OTHER INFORMATION: n equals a,t,g, or c
, NAME/KEY: misc feature
, LOCATION: (161)
, OTHER INFORMATION: n equals a,t,g, or c
, NAME/KEY: misc feature
, LOCATION: (178)
, OTHER INFORMATION: n equals a,t,g, or c
, NAME/KEY: misc feature
, LOCATION: (187)
, OTHER INFORMATION: n equals a,t,g, or c
, NAME/KEY: misc feature
, LOCATION: (203)
, OTHER INFORMATION: n equals a,t,g, or c
, NAME/KEY: misc feature
, LOCATION: (209)
, OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-11086

```

Query Match.	7.1%;	Score 146.6;	DB 9;	Length 225;
Best Local Similarity	91.9%;	Pred. No. 2.1e-33;		
Matches 171;	Conservative 0;	Mismatches 13;	Indels 2;	Gaps 2;
QY	851	GGGTATTGAAGATGCTGACAGAAACAGCAACTCAAACTGTGTGGCGACAGATCCACGGA	910	
Db	8	GTGTATTGAAGUGCTGAGAGAAACAGCAACTCAAACTGNTGTGGACAGATCCACGGA	67	
QY	911	AGGATTTGTTTAATATGGATGAAGATG-AAGAAACTGAACATGAACAAAGCGTGGTCCTG	969	
Db	68	AGGANTGNTTAATATGGATGAAGATGNAAGAAACTGAACATGAACAAAGCGTGGTCCTG	127	
QY	970	AACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCT-GGACATCTTGATCTCTTTG	1028	
Db	128	AACGGCTCTACCAAGATGGTGCATCCNNTAGCCGNGCGCCTGGGACATCCCGNATGTCTTTN	187	
QY	1029	GTTTTG 1034		
Db	188	GTTTTG 193		

RESULT 14  
 US-10-029-386-8224  
 ; Sequence 8224, Application US/10029386  
 ; Publication No. US20030194704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
 ; FILE REFERENCE: AEOMICA-X-2  
 ; CURRENT APPLICATION NUMBER: US/10/029,386  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 34288  
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 8224  
 ; LENGTH: 506  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO CHR16.3  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8  
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
 ; OTHER INFORMATION: SWISSPROT HIT: Q9NIV4, EVALUAE 7.10e-01  
 ; OTHER INFORMATION: EST HUMAN HIT: BG485267.1, EVALUAE 0.00e+00  
 ; OTHER INFORMATION: NT HIT: gi16164646, EVALUAE 5.00e-84  
 US-10-029-386-8224

	Query Match	6.8%;	Score 140.6;	DB 15;	Length 506;
	Best Local Similarity	96.9%;	Pred. No. 2.6e-31;		
	Matches 154;	Conservative 0;	Mismatches 4;	Indels 1;	Gaps 1;
Qy	1	ACAGAGCTGTGGCTGGAAGAGCTGGGCATCCGCGCTTGAGGCGCAGCGTGCGGTTAGT	60		
Db	214	ACAGAGCTGTGGCTGGAAGAGCTGGGCATCCGCGCTTGAGGCGCAGCGTGCGGTTAGT	273		
Qy	61	TCGCCCCAATGGCGGCACCGCTGTTTCACACG-CGTTTTCGGGAGATGCGGCGCGTTCG	119		
Db	274	TCGCCCCAATGGCGGCACCGCTGTTTCACACGTTGTTTTCGGGAGATGCGGCGCGTTCG	333		
Qy	120	TCCTCTGCAGTTAAGAAGCTGGGCGCGCTCAGAGACTGGG	158		
Db	334	TCCTCTGCAGTTCAAGACGCTGGGCGCGCTCAGAGACTGGG	372		

```

RESULT 15
US-10-029-386-21924
; Sequence 21924, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC A
; FILE REFERENCE: AEOmica-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21924
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HLA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P09133, EVALUJE 5.10e-01
; OTHER INFORMATION: NT HIT: G116164646, EVALUJE 1.00e-83
; OTHER INFORMATION: EST_HUMAN HIT: BG485267.1, EVALUJE 3.00e-76
US-10-029-386-21924

```

	Query Match	6.8%;	Score 139.6;	DB 15;	Length 342;
	Best Local Similarity	96.8%;	Pred. No. 3.9e-31;		
	Matches 153;	Conservative	0;	Mismatches 4;	Indels 1;
					Gaps 1;
QY	1	ACAGAGCTGTGCTCGAAGAGCTGGGCATCCGGCTGAGGCCACGGTCGCGTTAGT	60		
Db	185	ACAGAGCTGTGCTCGAAGAGCTGGGCATCCGGCTGAGGCCACGGTCGCGTTAGT	244		
QY	61	TCGGCCCAATGGCGGCACCCGCTGTTTACACAGC-CGTTTTCGGGAGATCGGCGCGCTTCG	119		
Db	245	TCGGCCCAATGGCGGCACCGCTGTTTACACAGTTGTTTTCGGGAGATCGGCGCGCTTCG	304		
QY	120	TCCTCTGCATTTAAGAGCTGGGCGCGTTCGAGACTGG	157		
Db	305	TCCTCTGCATTCAGACGCTGGGCGCGTTCGAGACTGG	342		

Search completed: July 18, 2004, 08:45:45  
Job time : 937 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 10:56:33 ; Search time 772 Seconds  
(without alignments)  
4113.969 Million cell updates/sec

Title: US-09-932-678-2

Perfect score: 3403  
Sequence: 1 MAAPLLHTRLPGDAAASSA.....RSPSSVGVSPVLYMQPSPL 651

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3190992 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US0932678/runat\_15072004.103606\_5312/app\_query.fasta\_1.839  
-DB=published Applications NA\_QPMT=fastap SUFFIX=rnpb-MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US0932678 @cgn1\_1.480 @runat\_15072004.103606\_5312  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:\*\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*\*
- 18: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	3403	100.0	2068	9	US-09-932-678-1	Sequence 1, Appli
2	3391	99.6	2040	9	US-09-952-013A-1	Sequence 1, Appli
3	2839	83.4	1770	16	US-10-094-466-7	Sequence 7, Appli
4	1124	33.0	1461	13	US-10-221-625-172	Sequence 172, App
5	1000	29.4	2493	16	US-10-094-749-544	Sequence 544, App
6	565.5	16.6	3878	15	US-10-128-714-207	Sequence 207, App
7	565.5	16.6	4139	15	US-10-128-714-5207	Sequence 5207, App
8	563	16.5	2139	15	US-10-128-714-6207	Sequence 6207, App
9	563	16.5	2139	15	US-10-128-714-7207	Sequence 7207, App
10	561.5	16.5	1878	15	US-10-128-714-1207	Sequence 1207, App
11	542.5	15.9	1830	15	US-10-128-714-2207	Sequence 2207, App
12	506.5	14.9	2702	13	US-10-424-599-77428	Sequence 77428, A
13	499.5	14.7	1671	15	US-10-032-585-6770	Sequence 6770, A
14	456.5	13.4	1815	9	US-09-938-842A-2204	Sequence 2204, Ap
15	456.5	13.4	1815	11	US-09-938-842A-2204	Sequence 2204, Ap
16	391.5	11.5	1293	16	US-10-320-797-2207	Sequence 2207, Ap
17	375	11.0	3902	16	US-10-320-797-207	Sequence 207, App
18	341	10.0	535	15	US-10-029-386-10613	Sequence 10613, A
19	337.5	9.9	1902	16	US-10-320-797-1207	Sequence 1207, Ap
20	335	9.8	1284	13	US-10-425-114-36547	Sequence 36547, A
21	333	9.8	535	15	US-10-029-386-12002	Sequence 12002, A
22	332.5	9.8	1103	17	US-10-437-963-99865	Sequence 99865, A
23	332	9.8	184	15	US-10-029-386-24313	Sequence 24313, A
24	327	9.6	184	15	US-10-029-386-25702	Sequence 25702, A
25	322	9.5	2103	9	US-09-925-301-192	Sequence 192, App
26	265.5	7.8	2591	17	US-10-437-963-99864	Sequence 99864, A
27	207	7.8	225	9	US-09-783-590-11086	Sequence 11086, A
28	202	5.9	140	9	US-09-728-445-420	Sequence 420, App
29	189	5.6	557	9	US-09-998-598-1177	Sequence 1177, Ap
30	180	5.3	414	9	US-09-864-761-4848	Sequence 4848, Ap
31	161	4.7	95	9	US-09-864-761-21583	Sequence 21583, A
32	135	4.0	598	15	US-10-029-386-7910	Sequence 7910, Ap
33	132.5	3.9	7884	13	US-10-424-599-113868	Sequence 113868, A
34	130.5	3.8	3255	13	US-10-424-599-754	Sequence 754, App
35	128	3.8	89	13	US-10-029-386-21610	Sequence 21610, A
36	127	3.7	1908	15	US-10-032-585-6636	Sequence 6636, Ap
37	120.5	3.5	2841	13	US-10-282-122A-18265	Sequence 18265, A
38	118	3.5	10116	17	US-10-437-963-98312	Sequence 98312, A
39	117.5	3.5	3018	13	US-10-425-114-12808	Sequence 12808, A
40	117	3.4	2628	15	US-10-032-585-6415	Sequence 6415, Ap
41	116.5	3.4	4371	17	US-10-437-963-92083	Sequence 92083, A
42	116	3.4	2483	13	US-10-424-599-12638	Sequence 12638, A
43	116	3.4	5164	17	US-10-437-963-14824	Sequence 14824, A
44	115	3.4	4398	13	US-10-282-122A-11087	Sequence 11087, A
45	112	3.3	506	15	US-10-029-386-8224	Sequence 8224, Ap

ALIGNMENTS

RESULT 1  
US-09-932-678-1  
; Sequence 1, Application US/09932678  
; Patent No. US2002090706A1  
; GENERAL INFORMATION:  
; APPLICANT: Reeder, Ronald H.  
; APPLICANT: Moorefield, Beth  
; APPLICANT: Greene, Elizabeth A.  
; TITLE OF INVENTION: HUMAN RXN3 AND COMPOSITIONS AND METHODS RELATING  
; TITLE OF INVENTION: THERETO  
; FILE REFERENCE: 14538A-005810US  
; CURRENT APPLICATION NUMBER: US/09/932.678  
; CURRENT FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: 60/225,893  
; PRIOR FILING DATE: 2000-08-16  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2068  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-932-678-1



## Alignment Scores:

Pred. No.: 0 Length: 2068  
Score: 3403.00 Matches: 651  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-932-678-2 (1-651) x US-09-932-678-1 (1-2068)

QY 1 MetAlaAProLeuLeuHisThrArgLeuProGlyAspAlaAlaSerSerAla 20  
DB 69 ATGGCGGCACCGCTGCTTACACACGGCTTTGCCGGAGATGGCGCTTCCTCTGCA 128  
QY 21 VallysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40  
DB 129 GTTAAGAAGCTGGGCGGCTCGAGGACTGGGATTTCAAAATATCGTGCAATAGAGAATGAC 188  
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60  
DB 189 TTTTTCATTTCTCCCCCAAGAAAACACTGTTGGTTTGGTGAACCTGTGACAGAAGTCTTG 248  
QY 61 LeuLysTyrLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
DB 249 CTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTTGTTGAAGAACCCAGCTGTTAGAT 308  
QY 81 ProAspIleLysAspAspGlnIleLeuAsnTrpLeuLeuGluPheArgSerSerIleMet 100  
DB 309 CAGACATAAAGGATGACAGATATCAACTCGCTGCTAGAAATTCGTTCTTCTATCATG 368  
QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleLeuLeuArgLeuProTrpLeuAsn 120  
DB 369 TACTTGCACAAAGACTTTGAGCACTTATCAGTATATATTAAGATTGCTTGGTTGAAT 428  
QY 121 ArgSerGlnThrValValGluGluThrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
DB 429 AGAAGTCAACAGTAGTAGAGAGATATTGGCTTTCTTGGTAAATCTTGTATCAGCACAG 488  
QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160  
DB 489 ACTGTTTTCCTCAGACCGGTCTCAGCATGATTTGCCATTTTGTGCTCCCGGAGTG 548  
QY 161 IleIleLysGluGlyAspValSerAspSerAspGluAspAspAsnLeuPro 180  
DB 549 ATCAATTAAGGAAGCGATGATGATGTTTCAGATCTGTGATGAAGAATGATATCTTCT 608  
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200  
DB 609 GCAAAATTTTGACACATGTACAGAGCCTTGCAAAATAATAGCAAGATATGTACCATCGACA 668  
QY 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220  
DB 669 CGTGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTTGTGGAAAAATCAGAGAGA 728  
QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
DB 729 ACACGTGAATGTACGTTTCAATTAACCTTACTAAGGATAGTGTATATTTTCCAACCTTGAG 788  
QY 241 HisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
DB 789 CATGAAATTCGAGGCTTATTTGAAAAAACTACTCAAGTTGGATGTGAATGCATCCCGG 848  
QY 261 GlnGlyIleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
DB 849 CAGGGTATTTGAGATGCTGAAGAAACAGCACTCAAACTTGTGGTGGACAGATTTCCAG 908  
QY 281 GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300  
DB 909 GAAGGATTTGTTAATATGATGAAGATGAAGAAACTGAACATGAACAAAGGCTGCTCT 968  
QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
DB 969 GAACGGCTGCACAGATGGTGCATCCTGTAGCCGAGCGCTGGACATCTCTGATGCTTTG 1028

QY 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340  
DB 1029 GTTTTGTCTCATGAAGGATGCTGCTATGTAGATGGTAAGTTGATAACGGCAAAACA 1088  
QY 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla 360  
DB 1089 AAGGATCTATATCGCGACTGATAAACAATCTTTGCAAACTCCTGTTGCCCAACCCATGCC 1148  
QY 361 SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380  
DB 1149 TCCTGCCATGTACAGCTTTTTCATGTTTACCTCTAGTTTCAAAATGGGATTCGCAGAG 1208  
QY 381 AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProIleLeuArg 400  
DB 1209 GCATTTTGGAAACATCTCTGGAAAAAATTCAGGACCCCAAGTAATCTCTGCCATCATCAGG 1268  
QY 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420  
DB 1269 CAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAAGAGCTAAATTTATCTCTTATT 1328  
QY 421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440  
DB 1329 ACTGTAAATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCCAG 1388  
QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460  
DB 1389 GATTCGGGAACAAAGCAATCTCGCATGTTGCTCTCCATGGACCAATTTTACTCAGCTGC 1448  
QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
DB 1449 CAACTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTGAGCGGAACCTGAAA 1508  
QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500  
DB 1509 GAAGGTTTTCAGATCTCTTCAGAGTCTGAATTTTGGACGAGATAGTATGAGCCAGCTAAAT 1568  
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr 520  
DB 1569 CCCCCTGAAGATTTGCTGCCCTCAGTGGTTAACTTTTTTGTGCAATCAAAATAGTAC 1628  
QY 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540  
DB 1629 CAGCTCGTCTTCTGCTACACCATCATTTGAGAGGAACAATCGCCAGATGCTGCCAGTCAAT 1688  
QY 541 ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560  
DB 1689 AGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTC 1748  
QY 561 ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580  
DB 1749 CCCTTTGATCCCTGTGCTGAGAGGTCAGAGAAATTCATTGATCTTATTTATCAGGTA 1808  
QY 581 TrpGluAspMetSerAlaGluLeuGlnPheLysLysProMetLysLysAspIle 600  
DB 1809 TGGGAAGACATGAGTGTCTGAAGAGCTCAGAGAGTTCAAGAAACCCATGAAAGGACATA 1868  
QY 601 ValGluAspGluAspAspPheLeuLysGlyValProGlnAsnAspThrValIle 620  
DB 1869 GTGGAAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCTCCAGATGATACCGTGAT 1928  
QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerValGlySer 640  
DB 1929 GGGATCACCAAGCTCTTTGACACGCAATTCGAGAGTCTTCAAGTAGTGTGGGCTCC 1988  
QY 641 ProProValLeuTyrMetGlnProSerProLeu 651  
DB 1989 CCACCCGTGTTGATCATGCACCCAGTCCCTC 2021

## RESULT 2

US-09-952-013A-1

; Sequence 1, Application US/09952013A

; Patent No. US20020146801A1



QY 541 ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560  
DB 1660 AGGAGTACCGCTGGAGGACTCAGTCAGATCTGCACAAACCGCTGCACACCTTCTTC 1719  
QY 561 ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580  
DB 1720 CCGTTTGATCCCTGTGTCTGAAGAGGTCAAGAAATTCATTGATCCCTATTATCAGGTG 1779  
QY 581 TrpGluAspMetSerAlaGluLeuGlnGluPheLysLysProMetLysLysAspIle 600  
DB 1780 TGGGAAGACATCAGTGCTGAAGAGTCAAGAGGATTCAGAAACCCATGAAAGGACATA 1839  
QY 601 ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620  
DB 1840 GTGGAAGATGAAGATGATGACTTCTGAAAGCGAAGTGCCTCAAGTAGTGGGCTCC 1899  
QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerValGlySer 640  
DB 1900 GGGATCACCAAGTCTCTTGACACGCAATTCGGAAGTCTTCAAGTAGTGGGCTCC 1959  
QY 641 ProProValLeuTyrMetGlnProSerProLeu 651  
DB 1960 CCACCGGTGTGTACATGCAACCCAGTCCCTC 1992

RESULT 3

US-10-094-466-7  
; Sequence 7, Application US/10094466  
; Publication No. US20030203363A1  
; GENERAL INFORMATION:  
; APPLICANT: SPYTEK et al.  
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM  
; TITLE OF INVENTION: AND METHODS OF USING  
; FILE REFERENCE: THE SAME  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US/10/094,466  
; PRIOR FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: 60/274,281  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/288,148  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/274,849  
; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: 60/275,235  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: 60/338,375  
; PRIOR FILING DATE: 2001-12-04  
; PRIOR APPLICATION NUMBER: 60/275,579  
; PRIOR FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 60/335,302  
; PRIOR FILING DATE: 2001-10-31  
; PRIOR APPLICATION NUMBER: 60/275,601  
; PRIOR FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 60/276,000  
; PRIOR FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 60/277,338  
; PRIOR FILING DATE: 2001-03-20  
; Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatIn 2.1  
; SEQ ID NO 7  
; LENGTH: 1770  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (40)..(1723)  
US-10-094-466-7

Alignment Scores:  
Fred. No.: 0 Length: 1770  
Score: 2839.00 Matches: 556  
Percent Similarity: 85.41% Conservative: 0

Best Local Similarity: 85.41% Mismatches: 5  
Query Match: 83.43% Indels: 90  
DB: 16 Gaps: 1  
US-09-932-678-2 (1-651) x US-10-094-466-7 (1-1770)  
QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAla 20  
DB 40 ATGGCGGCACCGCTCTCTCACACGCTTTCGGGAGATGCGCGCTTCGCTCTGCA 99  
QY 21 ValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40  
DB 100 GTTAAAGAGCTGGCGCGCTCGAGGACTGGGATTTCAAATATGCGTGCTAGAGAATGAC 159  
QY 41 PhePheAsnSerProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60  
DB 160 TTTTTCATTTCTCCCCCAAGAAAACTGTTCGGTTTGGTGGAACTGTGCAGAACTCTTG 219  
QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
DB 220 CTGAAGTACAAAAGGGTGAAACAATGACTTTGAGTTGTTGAAGAACCCAGCTGTAGAT 279  
QY 81 ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerIleMet 100  
DB 280 CCAGACATAAAGGATGACAGATCACTCAACTGGCTGTAGAAATTCGGTTCTTCTATCATG 339  
QY 101 TyrLeuThrLysAspPheGluGlnIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 120  
DB 340 TACTTGCACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCTTGGTTGAAT 399  
QY 121 ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
DB 400 AGAAGTCAACAGTAGTAGGAAGAGTATTGGCTTTTCTTGGTAATCTGTATCAGCACAG 459  
QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160  
DB 460 ACTGTTTTCTCAGACCGGTCTCAGCATGATTGCTTCCATTTTGTGCTCCCCGAGTG 519  
QY 161 IleIleLysGluGlyAspValSerAspSerAspAspGluAspAspAsnLeuPro 180  
DB 520 ATCATTAAGGAAGCGCATGTAGATGTTTCAGATTCTGATGATGAAGATGATATCTTCCT 579  
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200  
DB 580 GCAAAATTTGACATGTCAGAGCCCTGCAAAATAATAGCAAGATATGTACCATCGACA 639  
QY 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220  
DB 640 CCGTGGTTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTGAAAAATCAGAGAGA 699  
QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
DB 700 ACATGGAAATGTACGTTTCATAACTTACTAAGATTAGTGTATATTTTCCACCTTGAGG 759  
QY 241 HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
DB 760 CATGAAATCTGAGCTTATTATTGAAAACTACTCAAGCTGAGTGTGATGATCCCGG 819  
QY 261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
DB 820 CAGGGTATTGAAGATGCTGAAGAAACAGCAAAATCAAACTTGTGGTGGGAGATTCACG 879  
QY 281 GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300  
DB 880 GAAGGATTGTTTAATATG----- 897  
QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
DB 897 ----- 897  
QY 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340  
DB 897 ----- 897

```
QY 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla 360
Db 897 -----
QY 361 SerCysHisValGlnPheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380
Db 898 -----GGATTCGACAG 909
QY 381 AlaPheLeuGluHisLeuTriPlysLysLeuGlnAspProSerAsnProAlaIleIleArg 400
Db 910 GCATTTTGGACATCTTGGAAACCTTGCAGATCCAGTAACCTGCCATCATCAGG 969
QY 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420
Db 970 CAGGCTGCTGGAAATATATTGGAAGCTTTTGGCAAGAGCTAAATTTATTCTCTATT 1029
QY 421 ThrValLysSerCysLeuAspLeuValAsnTyrLeuHisIleTyrIleAsnGln 440
Db 1030 ACTGTAACCCATGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCCAG 1089
QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460
Db 1090 GATTCGGGAACAAGGCATCTCGCATGTGCTCTCCATGGACCAATTTTACTCAGCCTGC 1149
QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480
Db 1150 CAAGCTGTGTCTACACCTTTGTTTATAGACAACAAGCAGCTTTTGAGCGGAACCTGAAA 1209
QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500
Db 1210 GAAGTTTGCAGATTCCTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCGAGCTAAAT 1269
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr 520
Db 1270 CCCCTGAAGATTGCTGCCCTCAGTGTGTTAACTTTTGTGCAATCACAATTAAGTAC 1329
QY 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540
Db 1330 CAGCTCGTCTTCTGTCACCATCATTTGAGAGGAACAATCGCAGATGCTGCCAGTCATT 1389
QY 541 ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560
Db 1390 AGGAGTACCGCTGGAGGAGATCTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTC 1449
QY 561 ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580
Db 1450 CCCTTTGATCCCTGTGTGCTGAAGAGGTCAAGAAATTCATTGATCCTATTATTATCAGGTG 1509
QY 581 TrpGluAspMetSerAlaGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
Db 1510 TGGGAAGACATGAGTGTCTGAAGAGCTACAGAGTTCAAGAAACCCATGAAANAAGACATA 1569
QY 601 ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620
Db 1570 GTGGAAGATGAAGATGATCATCTTCTGAAGCGGAAGTCCCGCCAGAAATGATACCGTGATT 1629
QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerValGlySer 640
Db 1630 GGGATCACACCAAGCTCCTTTGACACGATTTCCGAAGTCTCTCAAGTAGTGTGGGCTCC 1689
QY 641 ProProValLeuTyrMetGlnProSerProLeu 651
Db 1690 CCACCCGTTGTATCATGCAACCCAGTCCCTC 1722
```

## RESULT 4

US-10-221-625-172

; Sequence 172, Application US/10221625

; Publication No. US20040033942A1

; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.

; APPLICANT: HILLMAN, Jennifer L.

; APPLICANT: BAUGHN, Mariah R.

```
; APPLICANT: YUE, Henry
; APPLICANT: LAL, Preeti
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: BANDMAN, Olga
; APPLICANT: TANG, Y. Tom
; APPLICANT: MATHUR, Preete
; APPLICANT: SHAH, Purvi
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: REDDI, Roopa
; TITLE OF INVENTION: TRANSCRIPTION FACTORS
; FILE REFERENCE: PF-0761 PCT
; CURRENT APPLICATION NUMBER: US/10/221,625
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PERL Program
; SEQ ID NO 172
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20040033942A1 2646274CBI
US-10-221-625-172
```

```
Alignment Scores:
Pred. No.: 2,74e-129 Length: 1461
Score: 1124.00 Matches: 256
Percent Similarity: 55.44% Conservative: 14
Best Local Similarity: 52.57% Mismatches: 52
Query Match: 33.03% Indels: 166
DB: 10 Gaps: 10
```

US-09-932-678-2 (1-651) x US-10-221-625-172 (1-1461)

```
QY 195 ArgTyrValProSerThrProTyrPheLeuMetProIleLeuValGluLysPheProPhe 214
Db 263 AATACAAAATTAGGGCACCGTGTCTTCTCATGCACTATGTTGGAAAATTTCCATT 322
QY 215 ValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerVal 234
Db 323 GTTCGAAAATCAGAGAACACTGGAATGTTACGTTTCATACTTACTAAGGATTAGTGT 382
QY 235 TyrPheProThrLeuArgHisGluIleGluLeuIleGluLysLeuLysLeu 254
Db 383 TATTTTCAACCTTGAGCATGAATTTCTGGAGCTTATTATTGAAAACCTACTCAAGCTG 442
QY 255 AspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCys 274
Db 443 GATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAGCAAACTCAACTTGT 502
QY 275 GlyGlyThrAspSerThrGluGlyPheAsnMetAspGluAspGluGluThrGluHis 294
Db 503 GGTGGGACAGATCCACGAGGATTTGTTAATATG----- 538
QY 295 GluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGluArgLeu 314
Db 538 ----- 538
QY 315 AspIleLeuMetSerSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLys 334
Db 538 ----- 538
QY 335 ValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeu 354
Db 538 ----- 538
QY 355 LeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPhe 374
Db 538 ----- 538
QY 375 LysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTyrLysLysLeuGlnAspProSer 394
```

```

Db      539  -----GGATTGCGAGGCAATTTTGGACATCTTTGGAAAACTTGCAGATCCAAGT 592
QY      395  AsnProAlaIleAArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAla 414
Db      593  AATCTGCGCATCATCAGGCAGGCTGCTGGAATTTATATGGAAGCTTTTGGCAAGAGCT 652
QY      415  LysPheIleProLeuIleThrValIysSerCysLeuAspLeuValAsnTrpLeuHis 434
Db      653  AAAATTATTCTCTTATTACTGTAAACCATCCCTAGATCTTTTGGTTAACTGGGTGCAC 712
QY      435  IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGly 454
Db      713  ATATACCTTAATAACAGANTTCGGGACAAAGGCATCTCGGAGTTGCTCTCCATGGA 772
QY      455  ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474
Db      773  CCATTTTACTCAGCCTGCCAAGCTGTGTCTACACCTTTGTTTATAGACACAAGCAGCTT 832
QY      475  LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIle 494
Db      833  TTGAGCGGAAACTGAAAGAGGTTTGCAGATATCCTCAGAGTCTGAATTTTGAGCGGATA 892
QY      495  ValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAla 514
Db      893  GTGATGAGCCAGCTAAATCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTGTCT 952
QY      515  AlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnAr 534
Db      953  GCAATCACCA---AAGATGAAGACTTGTG----- 977
QY      534  gGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAs 554
Db      978  -----GATATGGATGTGTGTATGGT-----TGCACAAC 1006
QY      554  nProLeuAspThrPhePheProPheAspProCysValLeuLys----- 568
Db      1007  AAT-ATCAATTATTTATACCACCTGACCGTGCACTTCAAAATGGTTAAGATGGTAAGT 1065
QY      569  -----ArgSerLysLysPheIleAspPr 576
Db      1066  TTTATGCTGTATTTTACCATAATAAAAAAATTTGATAGGGGAAAAACAGTCTGCCTCC 1125
QY      576  oileTyrGlnValTrpGlu-----AspMetSerAlaGluG 588
Db      1126  ACTTTTG---ATATGGAGCTGTAAACATCTTCACCTTCGGTCTCCCTTCTGCCCCACA 1182
QY      588  uLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluAspAspPh 608
Db      1183  TCTGGGCAAGCTAAGAAAGCCT----- 1204
QY      608  eLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrProSer----- 625
Db      1205  -----GTTGCTCTCTCTGCGCACAGCTGGAATTCATACCCCAAGGCCCTA 1254
QY      626  ---SerPheAspThrHisPheArgSerProSerSerSerValGlySerProProValle 644
Db      1255  GCCTTCCACCACGACCAACATTTCAATCCCATCCCATCCCATCCCATCCCATCCCATCC 1306
QY      644  uTyrMetGlnProSerPro 650
Db      1307  -----CATCCCATCCCATCC 1318
RESULT 5
US-10-094-749-544
; Sequence 544, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
```

```

; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 544
; LENGTH: 2493
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-544

Alignment Scores:
Pred. No.:      2,27e-113      Length:      2493
Score:          1000.00      Matches:    188
Percent Similarity: 97.45%      Conservative: 3
Best Local Similarity: 95.92%      Mismatches: 5
Query Match:    29.39%      Indels:    0
DB:             16          Gaps:      0

US-09-932-678-2 (1-651) x US-10-094-749-544 (1-2493)

QY      374  PheLysLeuGlyPheAlaGlnAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393
Db      703  TATCGACCGGAGTTCGACAGGCAATTTTGGAAACATCTTTGGAAAAATTTGACAGGATCCA 762
QY      394  SerAsnProAlaIleAArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413
Db      763  AGTAATCTGCGCATCATCAGGCAGGCTGCTGGAAATTTATTTGGAAGCTTTTGGCAAGA 822
QY      414  AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuValAsnTrpLeu 433
Db      823  GCTAAATTTATTCTCTTATTACTGTAAACCATGCTAGATCTTTTGGTTAACTGGCTG 882
QY      434  HisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHis 453
Db      883  CACATATACCTTAATAACAGGATTCGGGAACAAAGGCATTTGCANTGTGCTCTCCAT 942
QY      454  GlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln 473
Db      943  GGAACATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTATAGACACAAGCAG 1002
QY      474  LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArg 493
Db      1003  CTTTTCAGCGGAAACCTGAAAGAGGTTTGCAGATATCTTTCAGAGTCTGAATTTTGAGCGG 1062
QY      494  IleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePhe 513
Db      1063  ATAGTGATGAGCCAGCTAAATCCCTGAAGATTTGCTGCCCTCAGTGGTTAACTTTT 1122
QY      514  AlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn 533
Db      1123  GCTGCAATCACAAATAAGTACGAGCTCGTCTTCTGTACACCATCATCGAGAGGAAACAAT 1182
QY      534  ArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThr 553
```

```
Db 1183 CGCCAGATCGTCGAGTCATTAGGACTACCACTGGAGGAGACTCAGTCGAGACCTGCACA 1242
QY 554 AnProLeuAspThrPhePheProPheAspProCysValLeuLysArg 569
Db 1243 AACCCACTGGACACCTTCGTCCTTGTATCCCTGTCGTCGAGGAGG 1290
RESULT 6
US-10-128-714-207
; Sequence 207, Application US/10128714
; Publication NO. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Weng
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroskin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 207
; LENGTH: 3878
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-207
Alignment Scores:
Pred. No.: 1,05e-58 Length: 3878
Score: 565.50 Matches: 182
Percent Similarity: 44.08% Conservative: 138
Best Local Similarity: 25.07% Mismatches: 264
Query Match: 16.62% Indels: 142
DB: 15 Gaps: 24
US-09-932-678-2 (1-651) x US-10-128-714-207 (1-3878)
QY 13 AspAlaAlaSerSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyLeuSer 32
Db 854 AACAGCGCGGAGCTGCGGGGCTAAACACAAAGTTGGCGGACATGGATCTTCGCTGCA 913
QY 33 AsnMetArgAlaLeuGluAsnAspPhePheAsnSer-----ProProArgLysThrVal 50
Db 914 CCCTCGTCCTCGTCGTCGTCGATGGCAGCGGCGCCAGCCAGTCGCAAGGAGCGGTT 973
QY 51 Arg-----PheGlyGlyThrValThr 57
Db 974 CGGGTTCAATTTGATAAGGATGTGGAATGAGAGAAGCGCTTGGAGTGTCAAGAGAAA 1033
QY 58 GluValLeuLeuLysTyrLysLysGly----- 66
Db 1034 GAAATGCTGTGAGTACGAAAGAGCGCGCGCTGTTGTAAGAAAGTGGCGAGAGCT 1093
QY 67 -----GluThrAsnAspPheGluLeuLeuLysAsnGlnLeu 78
Db 1094 ATCCAGCGGCATGTCCTCGAACCGATAGTAGAGCTTACGATCGATATCAAGAGATTTC 1153
QY 79 ---LeuAspProAspLysLysAspGln----- 87
Db 1154 TCGGCAGATCTCGCCCTCTGGAGAGAGATGGCTTGCTCCGAGATTACCAACGCAC 1213
```

```
QY 88 -----IleLeuAsnTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAsp 105
Db 1214 ACTACATTGAAGCATCATCTCATGGGTCTTCTGTCCTCAATGTTGCGTCGCTGGATCGCAAT 1273
QY 106 PheGluGlnLeuIleSerIleLeuArgLeuProTrpLeuAsnArgSerGlnThrVal 125
Db 1274 TGCAAACGGCTTGGTCAATGCAGTCCTGCACAGTGAATGGCTCGGGCGGATGAGTCGTAT 1333
QY 126 ValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArg 145
Db 1334 ATCAAGTTGTACATTCAATTCGATTTCTAGTTAACTTGGCGAGCCAGGAAAGCTATCTCGG 1393
QY 146 ProCysLeuSerMetIleAlaSerHisPhe-----Val 156
Db 1394 GGGTCTTGAAGATGTTGGCCACAAATTTGGGGGAGATCCCAAAGGGAGCGGCAAGCTA 1453
QY 157 ProProArgValIleIleLysGluGlyAspValSerValSerAspSerAspGluAsp 176
Db 1454 CTGGCTACGCCGCCAGTCCAAAGCTGCAGAGATC----- 1486
QY 177 AspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyr 196
Db 1487 -----TATAGCAGAACTCATATGGCTCTACGTCTATGTGATGCGATTG 1528
QY 197 ValProSerThrProTrpPheLeuMetProIleLeuValGluLysPhePhePropheValArg 216
Db 1529 ATTCCTTCTGGTAGTGGACCCATCCCCATCTTGTCTATGCGAGTTTCCCTTCGATACG 1588
QY 217 LysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPhe 236
Db 1589 GATTCCGCAAGGCGCAACATTGCGTACACGCGGAATCTTCTAAAGTCAATTAGTACGCC 1648
QY 237 ProThrLeuArgHisGluIleLeuLeuIleLeuLysLeuLeuLysLeuVal 256
Db 1649 CCTGAGCTCCAAAGCGGACATCTAGCTCTCATCGGAGAAAGTCGTAGATCGATGTC 1708
QY 257 AsnAlaSerArgGlnGlyIleGluAspAlaGluThrAlaThrGlnThrCysGlyGly 276
Db 1709 CAGATT---CAGGTAGATATGGAAGACATCGAGGACGAGTGGCGGAGGATGTTCTCCAC 1765
QY 277 ThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluThrGluHisGluThr 296
Db 1766 GCAGTCTCGCTGAAGCAGCATCTTGGAGATGATCGAGATGATGAGTTGACAAATGCATCC 1825
QY 297 LysAlaGlyProGluArgLeuAsp-----GlnMetValHisProValAlaGlu--- 312
Db 1826 GTCGCTAGCGATGATTCCTTAGAGTCCGCGGAGTCCGCGAGTCAAGGCCATCAAAGACAAC 1885
QY 313 -----ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyr 330
Db 1886 ATACTCAATTTGATGCGATGATCGATTCGCTGTTTGAATACTACGCCCGCGCT---TTC 1942
QY 331 ValAspGlyLysValAspAsnGlyLysThrLys---AspLeuTyrArgAspLeuIleAsn 349
Db 1943 ACCACGGCGCATTTGGATGACAAGAGATGCTGCGATTGTTA-----CTTCTGAGC 1993
QY 350 IlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPhe 369
Db 1994 CATTTCCAAAGTATTATCTCTCCGACCTATCGCTCCGCGACTCTCAATTTCTCTCTC 2053
QY 370 TyrLeuCysSerPheLysLeu-----GlyPheAlaGluAlaPheLeuGluHis 385
Db 2054 CACTTCTCCAGTCGTCGCTATCTGCTGGATCGATTTCCCGCAACCTGTGTTCTAGCTC 2113
QY 386 LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaGlyAsn 405
Db 2114 ATCTTCAATAAGATGCAA-----CCTGGATCTCTACGCGAGTCCGCGAGGCC 2161
QY 406 TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuLeuThrValLysSerCys 425
Db 2162 TACCTCGCTAGCTTCTGTCCTGCGGACATCTCTGGCGAGGTTGTAAGAGATGTC 2221
```

Qy 426 LeuAspLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLys 445  
Db 2222 TTTGACCTC-----TTAGGCACCCATCTGAACAACCTCCGCTTGTGACTATGAA 2269  
Qy 446 AlaPheCys-----AspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla 462  
Db 2270 GCCACTGTGAGGGCCGTGACTTCGCGGTACGCTCTTCTACTCCACGGCAGGCA 2329  
Qy 463 ValPheTyrThrPheValPheArgHisLysGlnLeu-----LeuSerGlyAsn 478  
Db 2330 TTGCTGTATATCTTCTGCTTCGCTGGCGAGACCTGACGCGCAGCGATGAGGGCGAT 2389  
Qy 479 LeuLysGluGlyLeuGlnTyrLeu-----GlnSerLeuAsnPhe----- 491  
Db 2390 ACCCTGAACAATTTGATGAATGGAGCGAGAGACATCACGTTCCCTCCCTCGGTCAAG 2449  
Qy 492 -----GluArgValMetSerGlnLeuAsnProLeuLysLysLysCysLeuProSer 508  
Db 2450 GAGGTCCCTCACCAAGCAATCCATTCCAACTGAACCCCTCAAAGTATGCTCGCGGGG 2509  
Qy 509 ValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle 528  
Db 2510 ATTGTGTCTGAATTCGCCGTATGGCTCAGCACCTCAACTTATGTATGTTTTCAGCAT 2569  
Qy 529 IleGluArgAsnAsnArgGlnMetLeuProValIleArgSer----- 542  
Db 2570 CTCGAGACGAACAAAGACTGCGCGTCTCTCGTCCGAGGATCTTCACCCATGGGAGAC 2629  
Qy 543 -----ThrAlaGlyGlyAspSerValGlnIleCysThr 553  
Db 2630 CCGCGCTTCAGCCATGTCGAGCGAGAACTCGTGGCGCGATGACCTGGGTTAT----- 2683  
Qy 554 AsnProLeuAspThrPhePheProPheAspProCysValLeuLysArgSerLysPhe 573  
Db 2684 ---CAGCTTGATGCTTATTTTCCCTTTGACCCATACAGCTACTCTCGAGCGCGCTTGG 2740  
Qy 574 IleAspProIleTyrGlnValTrpGluAspMetSerAlaGluGluLeuGlnPheLys 593  
Db 2741 CTGGAGGATGATTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2794  
Qy 594 LysProMetLysLysAspIleValGluAspGluAspAspPheLeuLysGlyGluVal 613  
Db 2795 -----GAGGCGATTTCGACACAGTACCTCGACGGGAGCGAA 2830  
Qy 614 ProGlnAsnAspThrValIleGlyIleThrProSerSerPheAspThrHisPheArgSe 633  
Db 2831 TCCGAAGGAGAC---CTGAGCGAGCTCACTGAGACCGAGCGAGACGACTGAATACTCTGA 2887  
Qy 633 rProSerSerVal 638  
Db 2888 ACCAAAACCTTGTT 2903

## RESULT 7

US-10-128-714-5207  
; Sequence 5207, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Wenqi  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Eroshkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; FILE OF INVENTION: Methods of Use  
; FILE REFERENCE: 10182-018-999  
; CURRENT FILING DATE: 2002-04-23  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890

; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5207  
; LENGTH: 4139  
; TYPE: DNA  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-5207

## Alignment Scores:

Pred. No.:	1,18e-58	Length:	4139
Score:	565.50	Matches:	182
Percent Similarity:	44.08%	Conservative:	138
Best Local Similarity:	25.07%	Mismatches:	264
Query Match:	16.62%	Indels:	142
DB:	15	Gaps:	24

US-09-932-678-2 (1-651) x US-10-128-714-5207 (1-4139)

Qy 13 AspAlaAlaAspSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSer 32  
Db 1115 AACAGCGCGGAGCTGCGGGGCTAAACACAAGTTGGCGACATGGATCTTTCGCTGTCA 1174  
Qy 33 AsnMetArgAlaLeuGluAsnAspPhePheAsnSer-----ProProArgLysThrVal 50  
Db 1175 CCTCGTCTCGTCTCGCTTCCGATGGCAGCGATGGCAGCCCGAGTCCAAGGAAGCGGGT 1234  
Qy 51 Arg-----PheGlyGlyThrValThr 57  
Db 1235 CGGGTTCAATTGATAAGGATGTGGAATGAGAGAAGCTCTTGGAGTGTCAAGAGAAA 1294  
Qy 58 GluValLeuLeuLysTyrLysLysGly----- 66  
Db 1295 GAAATGGCTGTGATACGACGAAAGAGCGCGCGCTTCTGTTGTAAGAACTGCGCAGAGCT 1354  
Qy 67 -----GluThrAsnAspPheGluLeuLeuLysAsnGlnLeu 78  
Db 1355 ATCCAGCGCATGTCTCTGGAACGGATAGTAGAGCCCTACGATCGTATCAAGAGATTTTC 1414  
Qy 79 ---LeuAspProAspIleLysAspGln----- 87  
Db 1415 TCGGCAGATCTTCGCGTCTGGAGAGAGATGGCTTCCCTCCCGAAGATTTTACCACGCGAC 1474  
Qy 88 -----IleIleAsnTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAsp 105  
Db 1475 ACTACATTGAAGCATCATCTCATGGGTCTTCTGTCCAATGTTGCTGCTGCTGCTGCTGCT 1534  
Qy 106 PheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrVal 125  
Db 1535 TGAACGGCTTGGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1594  
Qy 126 ValGluGluTrpLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArg 145  
Db 1595 ATCAAGTTGTACATTTCGATTTCTAGGTAACTTTGCGGCGAGCCCGGAGAGCTATCTCGG 1654  
Qy 146 ProCysLeuSerMetIleAlaSerHisPhe-----Val 156  
Db 1655 GCGGTCTTGAAGATGTTGGCAACAATTTTCGGGAGATCCCAAGGGGAGCGGCGAAGCTA 1714  
Qy 157 ProProArgValIleIleLysGluGlyAspValSerValSerAspSerAspGluAsp 176  
Db 1715 CTGGGTACCCCGGAGTCCAGCTGAGATC----- 1747  
Qy 177 AspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleAlaArgTyr 196  
Db 1748 -----TATAGCAGAACTCATATGGCTCTTACGTGATGATGATGATGATGATGAT 1789  
Qy 197 ValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheProPheValArg 216  
Db 197 ValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheProPheValArg 216



Db 1790 ATTCTCTTCGTAGTGGACCCATATCCCGCATCTTGTCTATCGAGTTTCCCTTCGTATACG 1849  
QY 217 LysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPhe 236  
Db 1850 GATTCGGCAAGGCCAACATTCGTACACGGCGAATCTCTAAAGGTCTATTAGTTACGCC 1909  
QY 237 ProThrLeuArgHisGluIleLeuGluLeuIleLeuGluLeuLeuLeuLeuLeuLeuVal 256  
Db 1910 CTTGAGCTCCAAAGCCGACCTTCTCTCATCACGGAGAAATCGTGAAGATCATGTC 1969  
QY 257 AsnAlaSerArgGlnGlyLeuGluAspAlaGluGluThrAlaThrGlnThrCysGlyGly 276  
Db 1970 CAGATT---CAGGTAGATATGAAGACATCGAGGACGAAGTCGGCGAGATGTTCTCCAC 2026  
QY 277 ThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThr 296  
Db 2027 GCAGTCTCGCTGAAGCCACGATGTTGGAGGATGACGAGATGAAGTTGACAATGCATCC 2086  
QY 297 LysAlaGlyProGluArgLeuAsp-----GlnMetValHisProValAlaGlu--- 312  
Db 2087 GTCGTACGCGATGATCCCTAGACGTGAGTCCCGCGAGTCAAGGCCATCAAGACACAC 2146  
QY 313 -----ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyr 330  
Db 2147 ATACTCAATTGGATGGATGATCGATTTCGTGTTGATTAATCTAGCCCGCGCT---TTC 2203  
QY 331 ValAspGlyLysValAspAsnGlyLysThrLys---AspLeuTyrArgAspIleLeuAsn 349  
Db 2204 ACCAGGCGCATTTGGATGACAGAGAGATGCACCTGGATTTA-----CTTCTGAGC 2254  
QY 350 IlePheAspLysLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPhe 369  
Db 2255 CATTTCCAAAGATTAATTCCTCCGACCTATCGCTCCCGCCACTCTCAATTCCTCTCTTC 2314  
QY 370 TyrLeuCysSerPheLysLeu-----GlyPheAlaGluAlaPheLeuGluHis 385  
Db 2315 CATCTCTCCAGTCGTCCCTATTCTGTGTGATCGATTGCGCACTCGTTCGTGAGTCC 2374  
QY 386 LeuTyrLysLysLeuGlnAspProSerAsnProAlaIleLeuArgGlnAlaAlaGlyAsn 405  
Db 2375 ATCTTCAATAAGATGCAA-----CCTGGAATCTACGCGAGTCCGCGACGCC 2422  
QY 406 TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys 425  
Db 2423 TACTCGCTAGCTTGTTGTCGCGGTGGTGGCAGATCTCTCGCGAGGTGTGAAGAGATGTC 2482  
QY 426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnGlnAspSerGlyThrLys 445  
Db 2483 TTTGACCTC-----TTAGGCACCATCTGAACAACCTCCGTCCTTGACTATGAA 2530  
QY 446 AlaPheCys-----AspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla 462  
Db 2531 GCCACCTGTGAGGCGCTGTATCTCGCGCTACGCTCTCTTCTACTCCACGCGCACAGCA 2590  
QY 463 ValPheTyrThrPheValPheArgHisLysGlnLeu-----LeuSerGlyAsn 478  
Db 2591 TTGCTGTATATCTTCTGCTTCGCTGGCGGACCTGACCGCCGAGCGATGGAGGCGCAT 2650  
QY 479 LeuLysGluGlyLeuGlnTyrLeu-----GlnSerLeuAsnPhe----- 491  
Db 2651 ACCCTGACCAAAATGATGAACCTGGAGCCAGAACATCATCGTTCCCTCCCTCGGTCAAG 2710  
QY 492 -----GluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer 508  
Db 2711 GAGGTCTCTTCAACCAAGCAATCCATTCCAAACTGAACCCCTCAAGATGTCGCGCGCG 2770  
QY 509 ValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle 528  
Db 2771 ATTGTGCTGTAATTCGCCCGGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2830  
QY 529 IleGluArgAsnAsnArgGlnMetLeuProValIleArgSer----- 542  
Db 2831 CTCGAGACGACAAAGACTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2890

QY 543 -----ThrAlaGlyGlyAspSerValGlnIleCysThr 553  
Db 2891 CCGCGCTTCAGCCATGTCGAGCGAGAACTCGTCCGCGGATGACTGGGTAT----- 2944  
QY 554 AsnProLeuAspThrPhePheProPheAspProCysValLeuLysArgSerLysLysPhe 573  
Db 2945 ---CAGCTTGATGCTATTTCCCTTTCACCATCACCAGTACTCTCGCAGCGCGCTTGG 3001  
QY 574 IleAspProIleTyrGlnValTrpIleAspMetSerAlaGluLeuGlnGluPheLys 593  
Db 3002 CTGGAGGATGATAGTCCACTGCGGTGATCCCTCGGTGACGACGACGAC----- 3055  
QY 594 LysProMetLysLysAspIleValGluAspGluAspAspPheLeuLysGlyGluVal 613  
Db 3056 -----GAGGCGGATTCGGACAGTACCTCGACGCGGAGCGAA 3091  
QY 614 ProGlnAsnAspThrValIleGlyIleThrProSerSerPheAsp-ThrHisPheArgSe 633  
Db 3092 TCCGAAGGAGAC--CTGAGCGAGCTCACTGAGACCGCAGACGACGACTGAATACTCTGA 3148  
QY 633 rProSerSerVal 638  
Db 3149 ACCAAAACCCCTTGT 3164

RESULT 8  
US-10-128-714-6207  
; Sequence 6207, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Wenqi  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Ershkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6207  
; LENGTH: 2139  
; TYPE: DNA  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-6207

Alignment Scores:  
Pred. No.: 7,48e-59 Length: 2139  
Score: 563.00 Matches: 177  
Percent Similarity: 44.03% Conservative: 133  
Best Local Similarity: 25.14% Mismatches: 254  
Query Match: 16.54% Indels: 140  
DB: 15 Gaps: 23

US-09-932-678-2 (1-651) x US-10-128-714-6207 (1-2139)

QY 13 AspAlaAlaAlaSerSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSer 32  
Db 115 AACAGCGCGCAGCTCGCGGGCTAAACACAGTTGGCCGACATGGATCTTTCGCTGCA 174



; APPLICANT: Hu, Wenci  
 ; APPLICANT: Tishkoff, Daniel  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Eroshkin, Alexey M  
 ; APPLICANT: Lemieux, Sebastien M  
 ; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
 ; TITLE OF INVENTION: Methods of Use  
 ; FILE REFERENCE: 10182-018-999  
 ; CURRENT APPLICATION NUMBER: US/10/128,714  
 ; PRIOR FILING DATE: 2002-04-23  
 ; PRIOR APPLICATION NUMBER: US 60/285,697  
 ; PRIOR FILING DATE: 2001-04-23  
 ; PRIOR APPLICATION NUMBER: US 60/287,066  
 ; PRIOR FILING DATE: 2001-04-27  
 ; PRIOR APPLICATION NUMBER: US 60/295,890  
 ; PRIOR FILING DATE: 2001-06-05  
 ; PRIOR APPLICATION NUMBER: US 60/303,899  
 ; PRIOR FILING DATE: 2001-07-09  
 ; PRIOR APPLICATION NUMBER: US 60/316,362  
 ; PRIOR FILING DATE: 2001-08-31  
 ; NUMBER OF SEQ ID NOS: 8603  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 7207  
 ; LENGTH: 2139  
 ; TYPE: DNA  
 ; ORGANISM: Aspergillus fumigatus  
 ; US-10-128-714-7207

## Alignment Scores:

Pred. No.: 7,486-59 Length: 2139  
 Score: 563.00 Matches: 177  
 Percent Similarity: 44.03% Conservative: 133  
 Best Local Similarity: 25.14% Mismatches: 254  
 Query Match: 16.54% Indels: 140  
 DB: 15 Gaps: 23

US-09-932-678-2 (1-651) x US-10-128-714-7207 (1-2139)

146 ProCysLeuSerMetIleAlaSerHisPhe-----Val 156  
 655 GCGGCTTGAAGATGTTGGCCAACTTTCGGGGAGATCCCAAAGGGGACGGCAAGCTA 714  
 157 ProProArgValIleIleLysGluGlyAspValSerAspSerAspGluAsp 176  
 715 CCTGGCTACGCCCCAGTCCCAAGCTGCAGAGATC-----747  
 177 AspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleAlaArgTyr 196  
 748 -----TATAGCAGAACTCATATGGCTCTACGTCAATGTGATGACAGTTG 789  
 197 ValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheProPheValArg 216  
 790 ATTCTCTTGGTAGTGGGACCTATCCCACTTCTGTATGCAAGTTCCCTTCGATACG 849  
 217 LysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPhe 236  
 850 GATTCGCCAAAGGCCCAACATTGCGTACACGGCGAATCTTCTAAAGGTCAATTAGTACGCC 909  
 237 ProThrLeuArgHisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuVal 256  
 910 CCTGAGCTCCCAAGCGCATCTTAGCTCTCATCACGAGAAAGTCTGTAAGATCGATGTC 969  
 257 AsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGly 276  
 970 CAGATT---CAGGTAGATATGGAGACATCGAGGACGAAGTGGCGGAGATGTTCTCCAC 1026  
 277 ThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThr 296  
 1027 GCAGTCTCGCTCGAAGCCACGATGTTGGAGGATGACGAGAATGAAGTTGACAAATGCATCC 1086  
 297 LysAlaGlyProGluArgLeuAsp-----GlnMetValHisProValAlaGlu--- 312  
 1087 GTCGCTACGATGATTCCTAGCTGGAGTCCGCGGAGTCAAGCGCATCAAGCAAC 1146  
 313 -----ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyr 330  
 1147 ATACTCAAAATGGATGGCATGATCGATTCGCTGTTGTAATACGTAGCCCGCGCT---TTC 1203  
 331 ValAspGlyLysValAspAsnGlyLysThrLys---AspLeuTyrArgAspLeuIleAsn 349  
 1204 ACCACGGGCACATTGGATGACAAAGAGAAATGCATCGATT---CTTCTGAGC 1254  
 350 IlePheAspLysLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPhe 369  
 1255 CATTTCAAAAGTATATCTCTCCGACCTATCGCTCCGCGCACTCTCAATTCCTCTCTTC 1314  
 370 TyrLeuCysSerPheLysLeu-----GlyPheAlaGluAlaPheLeuGluHis 385  
 1315 CACTTCTCCAGTCGTCGCTATTCTGTGTGATCGATTTGCGCAACCTGTGTTCAGTCT 1374  
 386 LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsn 405  
 1375 ATCTTCAATGAAGTCAAA-----CCTGCAATCTCTACGCGAGTCCGCGACGCC 1422  
 406 TyrIleGlySerPheLeuAlaArgAlaLysPheIleProIleLeuThrValLysSerCys 425  
 1423 TACTCTCTAGTCTGTTGCGCGGTGGCGCACTCTCTGCGGAGGTGTTGAAGAGATGTC 1482  
 426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrIleAsnAsnGlnAspSerGlyThrLys 445  
 1483 TTGTGACCTC-----TTAGGCACCACTGAAACCTCCGCTCTTGACTATGAA 1530  
 446 AlaPheCys-----AspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla 462  
 1531 GCCACCTGTGAGGCCCTGATCTTGGCGGTCTAGCTCTTCTTACTTCCACGCGCACGCCA 1590  
 463 ValPheTyrThrPheValPheArgHisLysGlnLeu-----LeuSerGlyAsn 478  
 1591 TTGCTGTATATCTTCTGCTTCCGCTGCGGAGACCTGACGACCGCAGCGATGGAGGGCGAT 1650

Qy 479 LeuLysGluGlyLeuGlnTyrLeu-----GlnSerLeuAsnPhe----- 491  
Db 1651 ACCCTGAACAAATGATGAACATGAGCAGCAGAGACATCACGTTCCCTCCCTCGGTCAAG 1710  
Qy 492 -----GluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer 508  
Db 1711 GAGGTCCCTCACCAAGCAATCCATTCACAACTGAACCCCTCAAGATGATGCTCGCGGCG 1770  
Qy 509 ValValAsnPhePheAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle 528  
Db 1771 ATTGTGTCTGAATTCGCCGTATGCTCAGCACCTCAACTTCATGATGCTGTTTCAGCAAT 1830  
Qy 529 IleGluArgAsnAsnArgGlnMetLeuProValIleArgSer----- 542  
Db 1831 CTCGAGACGACAAAGAGCTGCGCGTCTCTCGTTCGGGAGCATCTCAGCCATGGCAGAC 1890  
Qy 543 -----ThrAlaGlyGlyAspSerValGlnIleCysThr 553  
Db 1891 CCGCGCTTCAGCCATGTGAGCGAGAAACTCGTGGCGGCGATGACCTGGGTTAT----- 1944  
Qy 554 AsnProLeuAspThrPhePheProPheAspProCysValLeuLysArgSerLysLysPhe 573  
Db 1945 ----CAGCTTGATGCTATTTTCCCTTTGACCCATACCACTACCTCGCAGCGCGCGTGG 2001  
Qy 574 IleAspProIleTyrGlnValTrpGluAspMetSerAlaGluGlnLeuGlnGluPheLys 593  
Db 2002 CTGGAGGATGATTACCTCACTGGCGTGTGTATCCCTGGCGGTGGACGACGAGAC----- 2055  
Qy 594 LysProMetLysLysAspIleValGluAspGluAspAspPheLeuLysGlyGluVal 613  
Db 2056 -----GAGGCGGATTGCGACAGTCACTCGACGCGGAGCGAA 2091  
Qy 614 ProGlnAsnAsp 617  
Db 2092 TCGAAGAGAGAC 2103

## RESULT 10

US-10-128-714-1207  
; Sequence 1207, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Weng  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Eroskin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; TITLE OF INVENTION: Methods of Use  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128, 714  
; PRIOR FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1207  
; LENGTH: 1878  
; TYPE: DNA  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-1207

Alignment Scores: 9.14e-59 Length: 1878  
Pred. No.: 561.50 Matches:  
Score:

Percent Similarity: 45.71% Conservative: 124  
Best Local Similarity: 26.69% Mismatches: 253  
Query Match: 16.50% Indels: 101  
DB: 15 Gaps: 21  
US-09-932-678-2 (1-651) x US-10-128-714-1207 (1-1878)  
Qy 26 AlaSerArgThrGlyLeSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerPro 45  
Db 10 GCGTCTTGAGTGGTCAAGAGAAATGCTGTGAGTACGGAAGAGCCGCCGCTG 69  
Qy 46 ProArgLysThrValArg-----PheGlyGlyThrValThrGluVal 59  
Db 70 GTTCTGGAAGATCGCAGAGCTATCCAGCGGCATGCTCTGGAACGAGTGTGAGGCC 129  
Qy 60 LeuLeuLysTyrLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeu 79  
Db 130 TACGATCGTATCAAGAGATTTTCTCGCAGCATCTCGCGCTCGAGCAAGATGCTTG 189  
Qy 80 AspPro---AspIleLysAspGlnIleIle---AsnTrpLeuLeuGluPheArgSer 97  
Db 190 CCTCCGAGATTACCAACGACACTACATTGAAGCATCATCTCATGGTCTTCTGTCC 249  
Qy 98 SerIleMetTyrLeuThrLysAspPheGluGlnLeuLeuSerIleLeuLeuArgLeuPro 117  
Db 250 AATGTTGCGTCTGATCGCAATTCGAACGCGTGTGTCATGCAATGTCAGTCTGCACAGTAA 309  
Qy 118 TrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGluVal 137  
Db 310 TGCGTCGGCGGAGTGTGATATCAAGTTGTACATTTCGATTCTAGTAACCTTGGC 369  
Qy 138 SerAlaGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPhe----- 155  
Db 370 GCAGCCCGAGGAGCTATCTCGGGCGGCTTTGAAGATGTTGGCCAAACAATTTCGGGAG 429  
Qy 156 -----ValProProArgValIleIleLysGlyGluGlyAspValAsp 168  
Db 430 ATCCCAAGGGGAGCGGCAAGCTACTGCTGCTACGCCCAAGCTCCAGCTCAGAGATC--- 486  
Qy 169 ValSerAspSerAspAspGluAspAsnLeuProAlaAsnPheAspThrCysHisArg 188  
Db 487 -----TATAGCAACTCATATG 504  
Qy 189 AlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProIleLeu 208  
Db 505 GCTCTACGTCATGTGATGAGTGTCTCTCTGTGAGTGGACCCCTATCCCACTCTTG 564  
Qy 209 ValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsn 228  
Db 565 TCTATGCACTTTCCTTCGATACGATTCGCAAGGCCCAACATTCGCTACACGCGGAAT 624  
Qy 229 LeuLeuArgIleSerValTyrPheProThrLeuArgHisGluLeuLeuIle 248  
Db 625 CTTCCTAAGGTCATTTAGTTACGCCCTGAGCTCCAGCCGACATTTCTAGCTCTCATCAG 684  
Qy 249 GluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGlu 268  
Db 685 GAGAAAGTCGTGAAGATCGATGTCAGATT---CAGGTAGATATGGAAGACATCGAGGAC 741  
Qy 269 ThrAlaThrGlnThrCysGlyThrAspSerThrGluGlyLeuPheAsnMetAspGlu 288  
Db 742 GAAGTGGGCGGAGGATGTTCTCCACGCGAGTCTCGCCCTGAAGCCACGATGTTGGAGGATGAC 801  
Qy 289 AspGluGluThrGlnHisGluThrLysAlaGlyProGluArgLeuAsp-----Gln 305  
Db 802 GAGAAAGTGAAGTGAACATGATCGTCTGCTAGCGATGATTCCTTAGACGTGGAGTCCCGG 861  
Qy 306 MetValHisProValAlaGlu-----ArgLeuAspIleLeuMetSerLeuValLeu 322  
Db 862 CGAGTCAAGGCCCATCAAGAGCAACATACTCAAAATGGATGGCATGCTGATTCGCTTT 921  
Qy 323 SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLys--- 341

```
Db 922 GAATACTAGCCCGCGCT---TTCACCGGCGCACTTGGATGACAGAGAAATGCAGTG 978
Qy 342 AspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSer 361
Db 979 GATTTA-----CTTCAGCAATTTCCAAAGTATTATCTTCGACCATCGCTCC 1029
Qy 362 CysHisValGlnPheMetPheTyrLeuCysSerPheLysLeu-----Gly 377
Db 1030 CGCCACTCTCAATCTCTCTCTCCACTCTCCAGTCGTCCGCTATTCTGGTGATCGA 1089
Qy 378 PheAlaGluAlaPheLeuGluHisLeuTyrLysLeuGlnAspProSerAsnProAla 397
Db 1090 TTTGCCGCAACCTGGTTACGCTCATCTTCATATAGATGCAR-----CTGGA 1137
Qy 398 IleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIle 417
Db 1138 ATCTTACGCGCATCGCGCAGCAGCTACCTCGTAGCTTCGTTGCCGTGGTGGCACATC 1197
Qy 418 ProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeu 437
Db 1198 TCTGGCGAGTGTAGAGAGTCTTTGACCTC-----TTAGGCACCCATCTG 1245
Qy 438 AsnAsnGlnAspSerGlyThrLysAlaPheCys-----AspValAlaLeuHisGly 454
Db 1246 AACAACTCTCGTCTTACTATGACGACCACTGTCCAGGCCCTGATCTTCGCGCTACCGT 1305
Qy 455 ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474
Db 1306 CTTTCTACTCCAGCGCACAGCATTTGTATATATCTTCTGCTTCGCGTGGCGAGACCTG 1365
Qy 475 -----LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeu-----GlnSer 488
Db 1366 ACGACCGCAGCGATGGAGGGGATACCCCTGAACAAATGTGTAACCTGAGCCAGAGAC 1425
Qy 489 LeuAsnPhe-----GluArgIleValMetSerGlnLeuAsn 500
Db 1426 ATCAGTTTCCCTCCCTCGTCAAGGAGTCTCTCACCAAGCAATCCATCCAACTGAC 1485
Qy 501 ProLeuIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyr 520
Db 1486 CCCCTCAAGATATGCTCGCGCGGATGTGTCTGAATTCGCCCGTATGGCTCAGCACCTC 1545
Qy 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540
Db 1546 AACTTCATGACGTTCGCAATTCGAGACGAAACAAAGACTGCGCGTCTCTCGTTC 1605
Qy 541 ArgSer-----ThrAlaGly 545
Db 1606 CGGAGCATCTCAGCATGGCAGACCCCGCTTCAGCCATGTGAGCGAGAAACTCGTGCC 1665
Qy 546 GlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCys 565
Db 1666 GGCATGACCTGGGTAT-----CAGCTTGATGCTATTTTCCCTTTGACCCATAC 1716
Qy 566 ValLeuLysArgSerLysPheIleAspProIleTyrGlnValTrpGluAspMetSer 585
Db 1717 CAGTACTCGCAGCCCGGTGGTGAGATGATTCAGTCCACTGGCGTGGTATCCCT 1776
Qy 586 AlaGluLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluAsp 605
Db 1777 GCGGTGACGACCAAGGAC-----GAGGGCGATTCG 1806
Qy 606 AspAspPheLeuLysGlyGluValProGlnAsnAsp 617
Db 1807 GACAGTACCTCGACGGGAGCAATCCGAAGGAGAC 1842
```

## RESULT 11

```
US-10-128-714-2207
; Sequence 2207, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
```

```
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US/10/128,714
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2207
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-2207
```

```
Alignment Scores:
Pred. No.: 2,07e-56 Length: 1830
Score: 542.50 Matches: 169
Percent Similarity: 44.79% Conservative: 123
Best Local Similarity: 25.92% Mismatches: 243
Query Match: 15.94% Indels: 117
DB: 15 Gaps: 21
```

US-09-932-678-2 (1-651) x US-10-128-714-2207 (1-1830)

```
Qy 26 AlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPheAsnSerPro 45
Db 10 GCGTCTGGAGTGTCAAGAGAAAGAAATGGCTGTGATGACGAAAGAGCGCCCGCTG 69
Qy 46 ProArgLysThrValArg-----PheGlyGlyThrValThrGluVal 59
Db 70 GTTCGTGAAGAGTGCAGCAGAGTATCCAGCGCATCTCTGCAACGGATAGTGAAGCC 129
Qy 60 LeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeu 79
Db 130 TACGATCGTATCAAGAGAGATTTTCTCGGCAGATCCTCGCGCTCTGGACGAAGATGGCTT 189
Qy 80 AspPro---AspIleLysAspAspGlnIleIle---AsnTrpLeuLeuGluPheArgSer 97
Db 190 CTTCCCGAAGATTACCAACGCACACTACATTTGAGCATCATCTCATGGGTCTTCTGTGCC 249
Qy 98 SerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPro 117
Db 250 AATGTTGCGTCGCTGGATCGCAATTGCAACGGCTTGTCATTCAGTCAGTCCTGCACAGTAA 309
Qy 118 TrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuVal 137
Db 310 TGGCTCGGCGGATGAGTCGTATATCAAGTTGTATCATTTCTAGTAACCTTTCG 369
Qy 138 SerAlaGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPhe----- 155
Db 370 GCAGCCAGGGAAGCTATCTCGGGCGGTCTTCAAGATGTGTGCCAACAAATTTGGGGAG 429
Qy 156 -----ValProArgValIleIleLysGluGlyAspValAsp 168
Db 430 ATCCCAAGGGGACGGGCAAGCTACCTGGCTACGCCCGCATCCCAAGCTGCAGAGATC--- 486
Qy 169 ValSerAspSerAspAspGluAspAsnLeuProAlaAsnPheAspThrCysHisArg 188
Db 487 -----TATAGCAGAACTCATATG 504
```

```
QY 189 AlaLeuGlnIleIleAlaArgTyrValProSerThrProTppPheLeuMetProIleLeu 208
DB 189 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 505 GCTCTACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 564
QY 209 ValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsn 228
DB 209 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 565 TCTATGCGAGTTCCCTTCGATACGATTCGCGAAGGCCAATTCGCGTACACGGCGAAT 624
QY 229 LeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIle 248
DB 229 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 625 CTTCTAAAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 684
QY 249 GluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGlu 268
DB 249 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 685 GAGAAAGTCGTGAAGATCGATGTCAGATT---CAGGTAGATATGGAACATCGAGGAC 741
QY 269 ThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGlu 288
DB 269 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 742 GAAGTGGCGAGGATGTTCTCCACGAGTCTCGCTGMAAGCCACGATGTTGGAGGATGAC 801
QY 289 AspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAsp-----Gln 305
DB 289 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 802 GAGAATGAATGATGACAAATCATCTCGCTAGCGATGATTCCTAGACGTGGAGTCCCGG 861
QY 306 MetValHisProValAlaGlu-----ArgLeuAspIleLeuMetSerLeuValLeu 322
DB 306 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 862 CGAGTCAGGCCATCAAGACCAACATACCTCAATTTGGATGGCATGATCGATTCGCTGTTT 921
QY 323 SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLys--- 341
DB 323 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 922 GAATACTAGCCCGCGCT---TTCACCACGGCGCATTTGGATGACACAGGAGAATGCAC 978
QY 342 AspLeuTyrArgAspLeuIleAsnLysPheAspLysLeuLeuLeuProThrHisAlaSer 361
DB 342 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 979 GATTTA-----CTTCTGAGCCATTTCCAAAGATATATCTTCCGACCTATCGCTC 1029
QY 362 CysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeu-----Gly 377
DB 362 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1030 CGCCACTCTCAATTCCTCTCTCCATCTTCCATCTCCAGTCGTCGCTATTCTGGTGGATCGA 1089
QY 378 PheAlaGluAlaPheLeuGluHisLeuTyrLysLysLeuGluAspProSerAsnProAla 397
DB 378 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1090 TTTGCGCGAACCTGTGTTCCAGCTCATCTTCAATAAGATGCAA-----CCTGGA 1137
QY 398 IleIleArgGlnAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIle 417
DB 398 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1138 ATCTACCGCAGTCCGCGACAGCCTACCTCGCTAGCTTCGTTGCTGCGCGTACGCT 1188
QY 418 ProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeu 437
DB 418 -----LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeu-----GlnSer 488
QY 438 AsnAsnGlnAspSerGlyThrLysAlaPheCys-----AspValAlaLeuHisGly 454
DB 438 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1198 AACAACTCCGCTTCTGACTATGAGCGACCTGTCGAGGCGCTGATCTCGCGCTACGCT 1257
QY 455 ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474
DB 455 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1258 CTTTCTACTCCACGCGACAGGATGCTGCTATATCTTCTGCTTTCGCTGGCGAGACCTG 1317
QY 475 -----LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeu-----GlnSer 488
DB 475 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1318 ACAGCCGCGAGTGGGGCGATACCCCTGGAACAAATTGATGAATGGAGCGCAGAGAC 1377
QY 489 LeuAsnPhe-----GluArgIleValMetSerGlnLeuAsn 500
DB 489 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1378 ATCAGCTTCCCTCCCTCGTCAAGAGGCTCTTCACCAAGCAATCCATTCCAACATGAAC 1437
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr 520
DB 501 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1438 CCCCTCAAGATATGCTCGCGCGGATGCTGTAATTCGCCGTATGCTGCTGCTGCTGCT 1497
QY 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540
```

```
DB 1498 AACTTCATGTACGTTCTCAGCATTTCTCGAGACGACAAAGAGCTGGCGTCTCTCTCTC 1557
QY 541 ArgSer-----ThrAlaGly 545
DB 1558 CGGAGCATCTCAGCCATGGCAGACCCGCGCTTCAGCCCATGTGACGCGAGAAATCGTGCC 1617
QY 546 GlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCys 565
DB 1618 GCGCATGACCTGGGGTTAT-----CAGCTTGATGCTATTTTCCCTTTGACCCATAC 1668
QY 566 ValLeuLysArgSerLysPheIleAspProIleTyrGlnValTrpGluAspMetSer 585
DB 1669 CAGCTACCTCGCAGCCGCGTGGCTGGAGGATGATTACGTCCTGCGGTGATCCCT 1728
QY 586 AlaGluGluLeuGlnGluPheLysPheMetLysLysAspIleValGluAspGluAsp 605
DB 1729 GCGGTGGAGCAGCAGGAC-----GAGGCGGATTCG 1758
QY 606 AspAspPheLeuLysGlyGluValProGlnAsnAsp 617
DB 1759 GACAGTGACCTCGAGGAGCGAATCCGAGGAGAC 1794
RESULT 12
US-10-424-599-77428
; Sequence 77428, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: SOY Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 77428
; LENGTH: 2702
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_40935C.1
US-10-424-599-77428
Alignment Scores:
Pred. No.: 1.32e-51 Length: 2702
Score: 506.50 Matches: 178
Percent Similarity: 43.62% Conservative: 123
Best Local Similarity: 25.80% Mismatches: 245
Query Match: 14.88% Indels: 145
DB: 13 Gaps: 25
US-09-932-678-2 (1-651) x US-10-424-599-77428 (1-2702)
QY 16 AlaSerSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArg 35
DB 228 GCGCTTTCTCTCTCTCTGAAACGCCCTCGTTGGCTACCAACATATATTAGCATCACT 287
QY 36 AlaLeuGluAsnAspPhePheAsnSerProProArgLysThrValArgPheGlyGlyThr 55
DB 288 GCTTC-GAACCTCTCTTTTCTCAC-----CGACGCGAGTTTCCGGTTTAAAGGAAGC 337
QY 56 ValThrGluValLeuLeu-----LysTyrLysLysGlyGluThrAsnAspPhe 71
DB 338 ATGGGAAGAGTTTCAGGTGAGTGAACAAGAGAAATCTCTCAATGGAGAACACGATTTC 397
QY 72 Glu-----LeuLeuLysAsnGlnLeuLeuAspProAsp 72
DB 398 AGTGATTTCGAGTTAGTTTCATCTCAGAGGTGTGCTTGTTCAGTGACACACGGGTGAT 457
QY 73 -----LeuLeuLysAsnGlnLeuLeuAspProAsp 82
```

Db 458 CGAGAGATTACGATGAGTCGTCGTTATTTGACCCGAGAGCATCTTAATCTGAT 517  
QY 83 IleLysAspGlnIleLeuSerLeuLeuGluPheArgSerIleMetTyrLeu 102  
Db 518 ---GAGGTGCTATCTGTGACGACTTGAAGCGCTTCTGGACGAGTGCCTATATA 574  
QY 103 ---ThrLysAspPheGluGlnLeuIleSerIleLeuLeuArgLeuProTrpLeuAsnArg 121  
Db 575 GATTCTGTTCATCATGAATCCCTGCTCTTGTGTTTCGAGATGAGCCTCTGGAAATTAT 634  
QY 122 SerGlnThrValValGluLeuValPheLeuGlnIlePheLeuGlnValSerAlaGlnThr 141  
Db 635 GGAATTGATATATTGATGACCTGCTGGAACATCATCATATCTTGTGCTCTTCCAAATGA 694  
QY 142 ValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProArgValIle 161  
Db 695 AAATACATCGATGTTGTTGGAGATGCTGTGAAAATTTGTGCGGCTTTTACCTC 754  
QY 162 IleLysGluGlyValAspValSerAspAspGluAspAspAsnLeuProAla 181  
Db 755 TTA-----GATTCTGTGAGCGAGGAGAACGGAATTGAAAGA 790  
QY 182 Asn-----PheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValPro 198  
Db 791 AAAAATAAAGTTCTGTACGGGTACATGCAGCATTTGAAGAATAGCTGATTTGGTGCCT 850  
QY 199 SerThrProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSer 218  
Db 851 CTGACACCTTCGCGACTACTCCATAGTATTCAGAAAATGCGGAGTGTCTTCAGTAAG 910  
QY 219 GluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIle-----SerValTyrPhe 236  
Db 911 GAAACTGAGATTGTCATGATGTCAGAAATATGTTGAGCTGGAGAGTGTGCACTTGGGA 970  
QY 237 ProThrLeuArgHisGluIleLeuGluLeuIleLeuLysLeuLeuLysLeuAspVal 256  
Db 971 GAAACTGTTGTTCTTACCATGCTGCTCTCTGTGGATAGTGTCTGATAGTGTGATG 1030  
QY 257 AsnAlaSerArgGlnGlyIle-----GluAspAlaGluGluThrAlaThrGlnThrCys 274  
Db 1031 GAAATTTGGATGGATGGATGCTGCAAGACATGCT----- 1066  
QY 275 GlyGlyThrAspSerThrGluGluLeuPheAsnMetAspGluAspGluGluThrGluHis 294  
Db 1067 -----AAAGGCATATTTGAAATGAGATGATAGAGATGTTACTGCAATT 1108  
QY 295 GluThrLysAlaGlyProGluArgLeuAspGlnMetVal----- 307  
Db 1109 GCAGATGAG-----GATGAGATTTGTCATAGTATGCTCGTCAGAGTTATTAATAGA 1162  
QY 308 -----HisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu 322  
Db 1163 AAAAATTTGCAAGCAACTGGTTGTTGAGAAATAGATAGCTTAATG-----GTGCTG 1216  
QY 323 SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAsp 342  
Db 1217 GCATTTTGT-----CACCTTGAATCCTGTCAAAGCAGTGGCCGCTTGTGACTGAG 1264  
QY 343 LeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuProThrHisAlaSerCys 362  
Db 1265 GTATTCGATACTCTACTAGTCTCATTTTCAGAAACCGTATGTGATGCTACAGTCAAAA 1324  
QY 363 HisValGlnPheMetPheTyrLeuCysSer----- 373  
Db 1325 TTTACCCAGTATGATGTTCTATGATGATGTCACATGGACCTGAGGATGTGTGTGAAA 1384  
QY 374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393  
Db 1385 TTTGCCATGTTCTTCTGCTGATGTTT-----GGATGTGAT 1420  
QY 394 SerAsnProAlaIleLeuArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413

Db 1421 GTTAATCCCTTATTACTAGGATGAGTGTGCTTATCTTCTAGTATTATTGACTCGT 1480  
QY 414 AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeu 433  
Db 1481 GCAAGATTCTTCTTACGCTCATGTGTCATAACATCATCAAAAGCTTGGTTGATGCTGT 1540  
QY 434 HisIleTyr-----LeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAla 451  
Db 1541 TATGCATATTTGCAAGTTAGCTGATTTTGATATGAACCCAGCAGCA----- 1585  
QY 452 LeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrPheValPheArgHis 471  
Db 1586 ---CATCAAGTTTTTATTTCTGGTCCAGGCAATTATATATTCTGTGTTTCCGATG 1642  
QY 472 LysGlnLeuSer---GlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsn 490  
Db 1643 AGATCTTTGATGGATATCTCTTAGGCTTAAA-----TTGCAGCTGCTTAAT 1687  
QY 491 -----PheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer 508  
Db 1688 ATGCCCATGGAGGCAATCTTGAACATAAATGATGCTCTTGAAGGTGGTTTGCTACT 1747  
QY 509 ValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle 528  
Db 1748 GTAGTAGTGGAAATTTCTTAGACAGGCAAGGCTGCTCAACTTTTTCATGCGCGTCAGATCA 1807  
QY 529 IleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAlaGlyCysSer 548  
Db 1808 TTTGTTTTGATGATATGCTGTGAGTCTGATCTATCTTCAAGGCTTTTGGCGGATTCATAGA 1867  
QY 549 ValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCysValLeuLys 568  
Db 1868 -----CTTGACATGTTTTTCCCATTTGATCCATGTTTGTGTGAAG 1906  
QY 569 ArgSerLysLysPheIleAspProIleTyrGlnValTrp----- 581  
Db 1907 AAAAGTGAAGCTACATAAGGCCACATTTTGTGCTGTCCTCAAAAGTCAGAACTACATAT 1966  
QY 582 GluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleVal 601  
Db 1967 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2005  
QY 602 GluAspGluAspAspAspPheLeu-----LysGlyGluValProGlnAsnAspThr 618  
Db 2006 GAATTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2065  
QY 619 ValIleGlyIleThrProSerSerPheAsp 628  
Db 2066 ATGCTGAGCGTTGAAGGTTTTTGTGATTTTAAAC 2095  
RESULT 13  
US-10-032-585-6770  
; Sequence 6770, Application US/10032585  
; Publication No. US20030180953A1  
; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/032,585  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6770  
; LENGTH: 1671  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-10-032-585-6770  
Alignment Scores: 4.19e-51 Length: 1671  
Pred. No.:

Score:	49950	Matches:	155
Percent Similarity:	46.61%	Conservative:	98
Best Local Similarity:	28.62%	Mismatches:	219
Query Match:	14.68%	Indels:	72
DB:	15	Gaps:	15

US-09-932-678-2 (1-651) x US-10-032-585-6770 (1-1671)

Qy	81	ProAspIleLysAspAspGlnIleLeuAseTrpLeuLeuGluPheArgSerSerIleMet	100
Db	192	CCCCGAATCAATCACACTCCCAATTTCCAGATCGTCTGAA--CGAGCTCGTGTCAAA	248
Qy	101	TyrLeuThr-----LysaspPheGluGln-LeuIleSerIleLeuAArgLeuPr	117
Db	249	TACTCAGCGTTGGACAACACACCGCATGCGAACCTCATCTTTGCCATTTTGAGATACAA	308
Qy	117	oTrpLeuAsnArgSerGlnThr-----ValValGluGluTyrLeuAlaph	132
Db	309	CTGGCTAGACATAGCCGATACCGATGGGTACCACCGTTTCGCTGAATCGTATTCCTCAAT	368
Qy	132	eLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMetIleAl	152
Db	369	CCTCGTTGCATTTGGTGTGCGCGCTCCCAAAATACCTCCCAAGTGCACAAACAACTAT	428
Qy	152	aSerHisPheValProProArgValIleLeuLysGluGlyAspValAspValSerAspSe	172
Db	429	CCAGGAATTTGACTCGGCCA-----	447
Qy	172	rAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIl	192
Db	448	-----AATACGACCCACCACCGAGAAATACTACCAA	479
Qy	192	eIleAlaArgTyrValProSerThrProTrpPheLeuMetProIleLeuValGluLysPh	212
Db	480	AATCATCAAAATACGTTCCCATCAATCAACTCAATCCCACTCTTCGAAAAAAAATCT	539
Qy	212	ePro--PheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuAr	231
Db	540	CCCCACCAATTTTCGTCTCGCAAAAGGAACTTACCACTAGTCGAGACCTAGTCGCG	599
Qy	231	gIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleLeuLysLe	251
Db	600	CGTCTGGACTACACTCCAGAGTTTACAATTCAGTATATGGCAACTAATCATCGAGTCGAG	659
Qy	251	uLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaTh	271
Db	660	TATATAAATAGACGTTGAATCCAGAACGAA--CTTGACGACCTAGACAC-----	708
Qy	271	rGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGl	291
Db	709	-----GAAGAGATCGAGATTTGATTAGTGGCGAAGAGACTCCGA	749
Qy	291	uThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAl	311
Db	750	CAGCGAAGACGAAATGGGAGGTCCACCACCTCCATCAAAATCCCTCTG-----	798
Qy	311	aGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrVa	331
Db	799	-TCCAAATCGATACATATATCAGCACCCCTCTTGACTACACCGCCCACTCATCTCCAT	857
Qy	331	lAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePh	351
Db	858	CGAA--GAACCTAAACACGGAAACGGTGTCACTCTTCAACACCATCAACTCCCTCTT	914
Qy	351	eAspLysLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLe	371
Db	915	CAAAAGCCACATCTTGGCCCAACCACTTACCAAAACTGATCCAGTTTATCTTATTCACAT	974
Qy	371	uCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHiIleuTrpLysLysLeuGl	391
Db	975	CACCCAGTTTCCAACTCGAAATTCGGCAGCATCACTTGTGTATGCTCATCGATGTGGCTTT	1034
Qy	391	nAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLe	411

Db	1035	CAACCCCGCGAAATCTTGAAAAAGCGTCTCAAGGCAATGCAGTACTTCTGCTGCATACAT	1094
Qy	411	uAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAs	431
Db	1095	CGCCAGAGCAAAACCTTTCCGCCACACAGGTCACTCTCATTTGTCAGTTATCTTTGTCGG	1154
Qy	431	nTrpLeuHisIleTyTrpLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAl	451
Db	1155	TTGGCTCAACAAGTTTATCATCGAAAGACAACAAGAGGTCGAGTCTAGTGGGACAAACA	1214
Qy	451	aLeuHisGlyPro-----PheTySerAlaCysGlnAlaValPheTyTrh	466
Db	1215	AGTCGGGGGTATGGAACGATTCAAACTATTTCAGCCGCACTTCAGGCATTTGCTTCAT	1274
Qy	466	rPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyTrLe	486
Db	1275	CTTCTGTTTCGGCCAC--GAACTACTCTCCACCACCGAAAAACGACTGGGAGTGCACCT	1331
Qy	486	uGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLe	506
Db	1332	CGACAAGTTC--TTCCAAAGAGCTATCTCTTACCAAGTTTAAACCCGCTAAGTTCGTGA	1388
Qy	506	uProSerValValAsnPhePheAlaIleThrAsnLysTyTrpGlnLeuValPheCysTy	526
Db	1389	CGAGACGCTGTTGTACATATTTCGCAAGTTGGCCACCAACTCAATGTCTGCTACTGCTA	1448
Qy	526	rThrIleIleGluArgAsnAsnArgGlnMetLeu-----ProValIleArgSerTh	543
Db	1449	CTCCATCATCGACACACATAACCGGAGAAATTTGGACACAAACGATTCCTTCGCCG	1508
Qy	543	rAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPheAs	563
Db	1509	AGTTGGCAACTTTAAACAGAAACAGAGTCTCTGGAGCTTCTGGAAACGCTACTTCCCA	1568
Qy	563	pProCysValLeuLysArgSerLysPheIleAspProIleTyGlnValTrp---G	582
Db	1569	TCCATTGGTCTTGGCCCATCGCAATCCATCGTCGCTGCCAAACTACATAGATGGTCAGA	1628
Qy	582	uAspMetSerAlaGluLeuGlnGluPheLysProMetLysLysAspIleValG	602
Db	1629	AGATACCCATCCGAGANA-----ACCGA	1652
Qy	602	uAspGluAspAsp	606
Db	1653	TAGTCAAGACCAT	1665

## RESULT 14

```

US-09-938-842A-2204
Sequence 2204, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRI1300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 2204
LENGTH: 1815
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-2204

```



Alignment Scores: 1.15e-45 Length: 1815  
Pred. No.: 456.50 Matches: 168  
Score: 42.56% Conservative: 98  
Percent Similarity: 26.88% Mismatches: 257  
Best Local Similarity: 13.41% Indels: 103  
Query Match: 9 Gaps: 23  
DB:

US-09-932-678-2 (1-651) x US-09-938-842A-2204 (1-1815)

QY 42 PheAsnSerProProArgLysThrValArgPheGlyThr-----Val 56  
|||  
Db 31 TTCAACTCAATGCAATGACACACGCTGAATATACCGACACAGATTGGTTGGCGTT 90  
|||  
QY 57 ThrGluValLeuLeuLysThrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsn 76  
|||  
Db 91 AGAAGAGGCCCTTGCATCTCTCCAAATGAGATACCGACGATTATAGTCAGCTTAAACA 150  
|||  
QY 77 GlnLeu---LeuAspProAspLysAspAsp-----GlnIleLeuAsnTrpLeu 92  
|||  
Db 151 GTGATGTGCTTACAGAGACGCTGATTTCGATGAGTGGCACAGCTTGAGACGCTTG 210  
|||  
QY 93 LeuGluPheArgSerSerIleMetThrLeuThr-----LysAspPheGluGln 108  
|||  
Db 211 AAAAGTCTATCAGTTTCGCTTGGATAGATTAGTTTCATCATATAAGATCTTCTTGA 270  
|||  
QY 109 LeuIleSerIleLeuArgLeuProTrpLeuAsnArgSerGln---ThrValValGlu 127  
|||  
Db 271 GCTATGAGCTG-----TGGTATCATAGTCACAGACCTAGTGTATGAT 315  
|||  
QY 128 GluThrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCys 147  
|||  
Db 316 GCATGTGGTGGACCTAATCATATCATCTGCGCTACTAGTGAAATATCTGATCCTTGT 375  
|||  
QY 148 LeuSerMetIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVal 167  
|||  
Db 376 TTGAATATGCTCGTAAGAAATTTTCAGTCAACTACTTTTAAAGCATAAAGTTTCACAA 435  
|||  
QY 168 AspValSerAspSerAspAspGluAspAspAspLeuProAlaAsnPheAspThrCysHis 187  
|||  
Db 436 CAATTAGTCAAAAAGATGACAGAA-----GTTTCATCCACGGGTGAT 477  
|||  
QY 188 ArgAlaLeuGlnIleAlaArgTyValProSerThrProTrpPheLeuMetProIle 207  
|||  
Db 478 GCAGCCCTTCACAGATTCTTATTAATTCCTCTGCTCCCTGGAATTTAGTGTCTATA 537  
|||  
QY 208 LeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyValHis 227  
|||  
Db 538 CTTGCGCAGAACATGCGGTAAATTTGACAAAAGGACCTTCTATAGTCACATATGTGAT 597  
|||  
QY 228 AsnLeuLeuArgIle-----SerValTyPheProThrLeuArgHisGluIleLeuGlu 245  
|||  
Db 598 AACCTGTGTAGGTGGAGATAGTCAATCGAGAAAGTTGTCGACGGTATCTTATG 657  
|||  
QY 246 LeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAsp 265  
|||  
Db 658 ATGTAATGAGAGGATGCTAGATTAGATTG----- 690  
|||  
QY 266 AlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsn 285  
|||  
Db 691 -----GTAAGTGGATGTGATGACTCTAATGAGGACGAGAGTTTCCAGTGGG 729  
|||  
QY 286 MetAspGluAspGluThrGluHisGluThrLysAlaGlyProGlu----- 301  
|||  
Db 730 ATGGAATTCGAGATGCGATTGAAGACACTGAATGAAGGACGAGAGTTTCCAGTGGG 789  
|||  
QY 302 ArgLeuAspGlnMetVal-----HisProValAlaGluArgLeuAspIleLeuMet 318  
|||  
Db 790 GCTCTAAAAACAATACTTCAGTGGAAATGTAGTCTCTGAACCTGTGGACAAATTGATG 849  
|||  
QY 319 SerLeuValLeuSerTyMetLysAspValCysTyValAspGlyLysValAspAsnGly 338  
|||

Db 850 GTCCTATTTTTCATCATCTAGAAATCC---TGTCAAAACCTCTGATCGTTTGGAT----- 900  
QY 339 LysThrLysAspLeuTyArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThr 358  
|||  
Db 901 -----GAGTATTTGAAATCCCTTTAAAGTCGTCGAGAACATATATCTTGAACACA 951  
|||  
QY 359 HisAlaSerCysHisValGlnPhePheMetPheTyLeuCysSerPheLys----- 375  
|||  
Db 952 TACAAAACAAAAATTTTCACAGTTTCTGAIGTGTCTATGCTGCTCCTAGATCTCTGAAAAT 1011  
|||  
QY 376 -----LeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393  
|||  
Db 1012 TGTGTTGTGAGATTGCGCAGTAAGCTGTGGAC-----ATATATCTCTCC 1056  
|||  
QY 394 SerAsnProAla---IleIleArgGlnAlaAlaGlyAsnTyIleGlySerPheLeuAla 412  
|||  
Db 1057 AGCAACACATGTCGACTTACTAGGATGAGTCAGTGGCTTATCTAGCTAGCTATTGTCT 1116  
|||  
QY 413 ArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrp 432  
|||  
Db 1117 CGTGAAGATTTTTCCTGCTCTTTTGTGGCTAGCATGTTGAAAGATTGGTGACGAG 1176  
|||  
QY 433 LeuHisIleTyLeuAsn-----AsnGlnAspSerGlyThrLysAlaPheCysAspVal 450  
|||  
Db 1177 TGTGCGGAATATGCGAACACATGCAATGATGATGTAAGCCACAGCA----- 1224  
|||  
QY 451 AlaLeuHisGlyProPheTySerAlaCysGlnAlaValPheTyThrPheValPheArg 470  
|||  
Db 1225 -----CATCAAGTCTTATTTGGAATGTCAGGCAATCTTGTATGTGTATGCTCCGA 1278  
|||  
QY 471 HisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyLeuGlnSer----- 488  
|||  
Db 1279 ATGAGATCCATAGTGGAG-----ATTCTCGCTTTCATTCGAGTTT 1320  
|||  
QY 489 LeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer 508  
|||  
Db 1321 AGATCATTTGGAGTCAATTTTATCTCACAAACTAAACCCACTACTGCTGTGCTTCCATCT 1380  
|||  
QY 509 ValValAsnPhePheAlaAlaIleThrAsnLysTyGlnLeuValPheCysTyThrIle 528  
|||  
Db 1381 GTAGTTTCCGAGTTCCTTAA--ACAAGCCAAAGCTGGTGTCTGTTCATTGTCTCAGAATC 1439  
|||  
QY 529 IleGlu--ArgAsnAsnArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSe 548  
|||  
Db 1440 CTTTCATTTTCGATGACCTACACGAGTCTGAGCTCTCTCGTCTTTTGGTGGCTTTGAAG 1499  
|||  
QY 548 rValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCysValLeuLys 568  
|||  
Db 1500 G-----CTTGACACATTTCTCCCGTTTGACCGCTGTGTTGAA 1538  
|||  
QY 568 sArgSerLysLysPheIleAspProIleTyGlnValTrp----- 581  
|||  
Db 1539 AATGCTCTAGCAGTACATCTCCCGAACTTCAACTCTTGTGCAATGGTGAACACGACTTA 1598  
|||  
QY 582 ----GluAspMetSerAlaGluLeuGlnGluPheLysLysProMetLysLysAspIle 600  
|||  
Db 1599 TGGAGAAGATGCTGACGAGAGCTTTGTGATGAGGTCAATAGTGAATGGAGATGCAGACAG 1658  
|||  
QY 600 eValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620  
|||  
Db 1659 TGGCGAGGAACCTTGACGATGACGTT-----GAACTTGTAGTGAATGAACACGAT 1709  
|||  
QY 620 eGlyIleThrPro 624  
|||  
Db 1710 GTCTACAATCCG 1722  
|||

## RESULT 15

US-09-938-842A-2204  
; Sequence 2204, Application US/09938842A  
; Publication No. US20040009476A9  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel

```

; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2204
; LENGTH: 1815
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2204

```

## Alignment Scores:

```

Pred. No.: 1,15e-45 Length: 1815
Score: 456.50 Matches: 168
Percent Similarity: 42.56% Conservative: 98
Best Local Similarity: 26.88% Mismatches: 257
Query Match: 13.41% Indels: 103
DB: 11 Gaps: 23

```

US-09-932-678-2 (1-651) x US-09-938-842A-2204 (1-1815)

```

QY 42 PheAsnSerProProArgLysThrValArgPheGlyGlyThr-----Val 56
   |||||
Db 31 TTCAACTCAATGCAATGACACACGCTGAATATATACGACACACAGATTGGTGTTCGCCGT 90
   |||||

QY 57 ThrGluValLeuLeuLysThrLysGlyGluThrAsnAspPheGluLeuLeuLysAsn 76
   |||||
Db 91 AGAAGGCCCTGCATCTCTCCAAATGAGATACCGACGATTAAGTACGCTTAACA 150
   |||||

QY 77 GlnLeu---LeuAspProAspLysAspAsp-----GlnleleAsnTrpLeu 92
   |||||
Db 151 GTGATGCTTACAGAAGCGCTGATTTTCGATGCGACGACAGCTTGACAGACTTG 210
   |||||

QY 93 LeuGluPheArgSerSerIleMetTyrLeuThr-----LysAspPheGluGln 108
   |||||
Db 211 AAAAGTCTATCAGTTTCCGTTGCTTGGATAGATTAGTTTCATCAATAAGATCTTGTAA 270
   |||||

QY 109 LeuIleSerIleLeuArgLeuProTrpLeuAsnArgSerGln---ThrValValGlu 127
   |||||
Db 271 GCTATGAGCTTG-----TGATATCATAGTCACAGACCTAGTGTATGAT 315
   |||||

QY 128 GluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCys 147
   |||||
Db 316 GCATTGGTGACCTAATCATATCATCTACTGCGCGTACTAGTGGAAATATCTGGATCCTGT 375
   |||||

QY 148 LeuSerMetIleAlaSerHisPheValProProArgValIleIleLeuGlyAspVal 167
   |||||
Db 376 TTGAATGCTGTAAGAAATTTTCGTAACCTCTTTTAAAGCATTAAGTTTCAACAAC 435
   |||||

QY 168 AspValSerAspSerAspGluAspAsnLeuProAlaAsnPheAspThrCysHis 187
   |||||
Db 436 CAATTAGTCAAAAGATGACAGAA-----GTTTCATCCACGGGTGCAT 477
   |||||

QY 188 ArgAlaLeuGlnIleLeuAlaArgTyrValProSerThrProTrpPheLeuMetProIle 207
   |||||
Db 478 GCAGCCCTTCAACAAGATTCTTATTTAATTCCTTCTGCTCCCTGGAAATTTAGTGTCTATA 537
   |||||

QY 208 LeuValGluLysPheProPheValArgLysSerGluArgThrGluCysTyrValHis 227
   |||||
Db 538 CTTGCGCAAGACATGCGTAAATTTGACAAAGGACCTCTATATGACATATGTGAT 597
   |||||

QY 228 AsnLeuLeuArgIle-----SerValTyrPheProThrLeuArgHisGluIleLeuGlu 245
   |||||
Db 598 AACCTGTTGAGTTGGAGATAGTCAATCGAGAAGTTGTTGGCAGCGGTGATCTTATG 657
   |||||

```

```

QY 246 LeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAsp 265
   |||||
Db 658 ATGTAATGGAGAGGATGCTAGATTGATTG----- 690
   |||||

QY 266 AlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsn 285
   |||||
Db 691 -----GTAAGTGGATGTGATGACTCTAATGGAGGATGTTTGTAT 729
   |||||

QY 286 MetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProGlu----- 301
   |||||
Db 730 ATGGAACCTTGAAGATGCGAGTTGAAAGCACTATGAATGAAGAGACGAGTTTCCACGTGGG 789
   |||||

QY 302 ArgLeuAspGlnMetVal-----HisProValAlaGluArgLeuAspIleLeuMet 318
   |||||
Db 790 GCTCTAAACAAATACTCTCAGGTGGAATGTAGTCTCTGAACTCTTGGACAAATGTATG 849
   |||||

QY 319 SerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGly 338
   |||||
Db 850 GTCTCTATTTTTCATCATCTAGATCC---TGTCAAAACTCTGATCGTTTGGAT----- 900
   |||||

QY 339 LysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThr 358
   |||||
Db 901 -----GAGGTATTTCAAAATCCTTTAAGTCGTGGAACAATAATATCTGAAACACA 951
   |||||

QY 359 HisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLys----- 375
   |||||
Db 952 TACAAAACAAATTTTTCAGTTTCTGATGTTCTATGATGCTCTACTAGATCCTGAAAT 1011
   |||||

QY 376 -----LeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393
   |||||
Db 1012 TGTGGTGTGAGATTTGCCAGTAAGCTGTTGGAC-----ATATATCTCTCC 1056
   |||||

QY 394 SerAsnProAla---IleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAla 412
   |||||
Db 1057 AGCAACACATGTCGACTTACTAGGATGAGTGAGTGGCTTACTAGTAGCTATTGTCT 1116
   |||||

QY 413 ArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrp 432
   |||||
Db 1117 CGTGGAAATTTTTCGCTCTCTTTTGGTGGTAGCATCTTGAAAAGATTGGTGGACGAG 1176
   |||||

QY 433 LeuHisIleTyrLeuAsn-----AsnGlnAspSerGlyThrLysAlaPheCysAspVal 450
   |||||
Db 1177 TGTGCGGAATATTGCGAAACATGCAATGATGATGTGAAGCCAGAACCA----- 1224
   |||||

QY 451 AlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArg 470
   |||||
Db 1225 -----CATCAAGTGTCTTATCTGATGTCAGGCAATCTGTATGTGCTATGCTCCGA 1278
   |||||

QY 471 HisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSer----- 488
   |||||
Db 1279 ATGAGATCCATAGTGGAG-----ATTCTCGCTTTCAATCGCAGTTT 1320
   |||||

QY 489 LeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSer 508
   |||||
Db 1321 AGATCATTTGAGTCAATTTTATCTCAAACTAAACCCACTACTGTTGCTCTCCATCT 1380
   |||||

QY 509 ValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle 528
   |||||
Db 1381 GTAGTTTCCAGTTCCTTAA-ACAAGCCAAAGCTGGTGTCTGTTCTTCTCTCAGATC 1439
   |||||

QY 529 IleGlu-ArgAsnAsnArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSe 548
   |||||
Db 1440 CTTTCATTTTCGATGACCTACACGAGCTGAGCTCTCTCGTCTTTTGGTGGCTTTGAAAG 1499
   |||||

QY 548 rValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCysValLeuLys 568
   |||||
Db 1500 G-----CTTGACATTTCTCCCGTTTACCCTGCTTGTGTTGAA 1538
   |||||

QY 568 sArgSerLysLysPheIleAspProIleTyrGlnValTrp----- 581
   |||||
Db 1539 AATGCTAGCAGCTACATCTCCCGCAACTTCACTTCTGTCATGTTGGTGAACACGCTTA 1598
   |||||

```

```
QY 582 ----GluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspI 600
Db      |||||
1599 TGGAGAGAGTGGTGACGAGAGCTTTGTGATGAGGTCATAGTGAATGGAGATGCACACAG 1658
QY 600 eValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValI 620
Db      ||::: |||||
1659 TCGGAGAGGAACCTGACGATGACGTT-----GAACTTGATAGTGAGATGAACACGAT 1709
QY 620 eGlyIleThrPro 624
Db      : |||||
1710 GTCTACAACTCCG 1722
```

Search completed: July 18, 2004, 14:02:24  
Job time : 818 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 08:30:12 ; Search time 126 Seconds  
(without alignments)  
2867.246 Million cell updates/sec

Title: US-09-932-678-2  
Perfect score: 3403  
Sequence: 1 MAAPLLHLRLPGDAASSA.....RSPSSVGVSPVLYMQSPSL 651

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delep 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US0932678/runat\_15072004\_103604\_5238/app.query.fasta\_1.839  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0932678 @CGN 1 1 69 @runat\_15072004\_103604\_5238 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	369	10.8	513	US-09-621-976-3086	Sequence 3086, Ap
2	138	4.1	263	US-09-313-294A-919	Sequence 919, App
3	132	3.9	262	US-09-313-294A-924	Sequence 924, App
4	111	3.3	4392	US-09-470-661A-3	Sequence 3, Appli
5	111	3.3	15413	US-09-981-282-3	Sequence 1, Appli
6	111	3.3	15424	US-09-981-282-2	Sequence 2, Appli
7	111	3.3	15424	US-10-143-186-2	Sequence 2, Appli
8	111	3.3	15450	US-09-470-661A-1	Sequence 1, Appli
9	106	3.1	15413	US-09-981-282-3	Sequence 3, Appli
10	106	3.1	15424	US-09-981-282-1	Sequence 1, Appli
11	106	3.1	15424	US-10-143-186-1	Sequence 1, Appli
12	104.5	3.1	580073	US-08-545-528D-1	Sequence 1, Appli

13	103.5	3.0	2115	1	US-08-329-681A-3	Sequence 3, Appli
14	103	3.0	8054	4	US-09-581-909-2	Sequence 2, Appli
15	102.5	3.0	1692	4	US-09-134-001C-335	Sequence 335, App
16	102	3.0	10968	3	US-09-228-246-1	Sequence 1, Appli
17	101	3.0	5475	2	US-08-680-327-1	Sequence 1, Appli
18	101	3.0	5475	3	US-09-228-246-3	Sequence 3, Appli
19	101	3.0	580073	4	US-08-545-528D-1	Sequence 1, Appli
20	100.5	3.0	4378	2	US-09-080-897-3	Sequence 3, Appli
21	100.5	3.0	4378	3	US-09-323-735-3	Sequence 3, Appli
22	100.5	3.0	4399	3	US-08-899-595-2	Sequence 2, Appli
23	100.5	3.0	6806	4	US-08-956-171E-472	Sequence 472, App
24	99.5	2.9	3024	4	US-09-957-005-8	Sequence 8, Appli
25	99.5	2.9	5361	3	US-08-973-462-2	Sequence 2, Appli
26	99.5	2.9	6152	3	US-08-973-462-1	Sequence 1, Appli
27	99	2.9	10323	4	US-09-280-428A-11	Sequence 11, Appli
28	98.5	2.9	2402	3	US-08-776-265-4	Sequence 4, Appli
29	98.5	2.9	2402	4	US-09-398-184-4	Sequence 4, Appli
30	98.5	2.9	4565	3	US-08-776-265-1	Sequence 1, Appli
31	98.5	2.9	4565	4	US-09-398-184-1	Sequence 1, Appli
32	98.5	2.9	5902	4	US-09-381-989-4	Sequence 4, Appli
33	98.5	2.9	89047	4	US-09-596-002-34	Sequence 34, Appli
34	97.5	2.9	4915	4	US-09-173-053-7	Sequence 7, Appli
35	97.5	2.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
36	97	2.9	1638	1	US-08-417-330A-13	Sequence 13, Appli
37	97	2.9	10968	2	US-08-680-327-2	Sequence 2, Appli
38	96.5	2.8	813	4	US-09-872-733A-19	Sequence 19, Appli
39	96.5	2.8	1360	4	US-09-082-649B-80	Sequence 80, Appli
40	96.5	2.8	1360	4	US-09-082-649B-81	Sequence 81, Appli
41	96.5	2.8	1738	4	US-09-574-873-9	Sequence 9, Appli
42	96.5	2.8	3418	2	US-08-944-916-12	Sequence 12, Appli
43	96.5	2.8	3547	4	US-09-340-798A-43	Sequence 43, Appli
44	96.5	2.8	3610	4	US-09-194-949A-1	Sequence 1, Appli
45	96.5	2.8	3987	4	US-09-082-649B-83	Sequence 83, Appli

ALIGNMENTS

RESULT 1  
US-09-621-976-3086  
; Sequence 3086, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET 054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; NUMBER FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 3086  
; LENGTH: 513  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 25..231  
US-09-621-976-3086

Alignment Scores:  
Pred. No.: 2,1e-36 Length: 513  
Score: 369.00 Matches: 75  
Percent Similarity: 88.64% Conservative: 3  
Best Local Similarity: 85.23% Mismatches: 5  
Query Match: 10.84% Indels: 6  
DB: 4 Gaps: 1

US-09-932-678-2 (1-651) x US-09-621-976-3086 (1-513)

Qy 528 l l e l e g l u a r g a s n a s n a r g i n m e t l e u p r o v a l l e a r g s e r t h r a l a g l y a s p 547  
Db 1 A T C A T T G A G G A G A C A A T C G C A G A T C T C C A G T C A T T A G G A T C A C C A C T G G A G G A C 60

Qy	548	SerValGlnIleCysThrAspProLeuAspThrPhePheProPheAspProCysValLeu	567
Db	61	TCAGTCGACACCTGCACAAACCCACTGGACACCTTCCTCCCTTGATCCCTGTGTGCTG	120
Qy	568	LysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspMetSerAlaGlu	587
Db	121	AAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTATGGGAAGCATGAGTGTGTGAA	180
Qy	588	GluLeuGlnGluPheLysLysProMetLysLysAspIleValGlu	602
Db	181	GAGCTACAGAGATTCAAGAAACCCATGAAGAAAGTT-GTTCGGGAATCATGTAGGCACA	239
Qy	603	AspGluAspAspPheLeuLys	610
Db	240	GATGAAGACGAACCTGCTTTGAAA	263

```

RESULT 2
US-09-313-294A-919
; Sequence 919, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 919
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Inbyte ID No. 6476212 700550063H1
; NAME/KEY: unsure
; LOCATION: 47-49, 262
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-919

```

[illegible]

```

; GENERAL INFORMATION:
; APPLICANT: Lalguindi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 924
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700550072H1
; US-09-313-294A-924

```

```

US-09-932-678-2 (1-651) x US-09-313-294A-924 (1-262)

Qy      406   TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValIysSerCys 425
         :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      1    TANGTTGGAGCTATCTGTCCCGGCACGGTTCATTCCGTAAATATAGTTCGTTCTGTA 60

Qy      426   LeuAspLeuIeuValAsnTrpLeuHisIleTyrIeuAsnAsnGln---AspSerGlyThr 444
         :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      61   CTCAAAAGGCTAGTGGAAGTGGTGCAATTATTATTGTGCCCATCAGTCGACAATGGGGTG 120

Qy      445   LysAlaPheCysAspValAlaLeuHiIsGlyProPheTyrSerAlaCysGlnAlaValPhe 464
         :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      121  ACAGCA-----AATCTTATAAGCACCAAGCTATTTTTATGTAGCTGCCAGCTGTGATG 174

Qy      465   TyrThrPheValPheArgHisLysGlnLeuLeuSer---GlyAsnLeuIysGluGlyLeu 483
         :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      175  CATGTCCTATGCTTTCGGTTAAGATCTATTATGGACTACCCAAATCTTAAACAACAACCTT 234

RESULT 4
US-09-470-661A-3
; Sequence 3, Application US/09470661A
; Patent No. 6500662
```

QY	307	lhIsProValAlaGluArgLeuAspIleLeuMetSerLeuValneuSerTyrMetLysAs	327
Db	1065	CCACCGGGCAGCG-----TTGAGTGGTGTCAACCACCGGCTTCATGAAAAA	1105
QY	327	pValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLe	347
Db	1110	GGCG-----TTTAACCTCGCCATCGCCCTCGGGAAAAA	1148
QY	347	uIleAsnIlePheAspLysLeuLeuProThrHisAlaSerCysHisValGlnPhePh	367
Db	1147	-----TTTAAGAGCTACAGACTCGGCTCTTAGCAGGTC-----	1188
QY	367	eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTr	387
Db	1183	-----CTTGAAGCTGATCTTGCATCTGCTG-----	1206
QY	387	pLysLysLeuGlnAspProSerAsnProAlaIleArgGlnAlaAlaGlyAsnTyrIle	407
Db	1207	-----GATCGATCCACACCTGCGCAATTGTCGCTGGTTTGCGCCAATCTTCT	1255
QY	407	eGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAs	427
Db	1254	TTATGAACTTGCTGTGCTGAAGAGCACCTACCGTCTGCTGAACCTGCTGCCATGA	1313
QY	427	pLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPh	447
Db	1314	CCATTGGTC-----ACGCAGTCCGGCGCAGTGACTAAGAG	1349
QY	447	eCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh	467
Db	1350	GGGTGGCTATCTGCTGGCGACCCGATCACTTCTGTCTCTAAACACCAATTACAGCTTGGT	1409
QY	467	eValPheArgHisLysGlnLeuLeuSer-----GlyAsnLeuLysGluGlyLeuGl	484
Db	1410	GATATATGCACACACATGTGTCTTAGTTACTTTAAAGTGGTCACCTCATGGCTTCT	1465
QY	484	nTyrLeuGln---SerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLys	503
Db	1470	GTTCCTACAGACCACTGAAGTTCGAGGAC-----ATGCTCAAGTCCACCCCTGAT	1522
QY	503	sIleCys-----LeuProSerValValAsnPhePhe-----	513
Db	1524	CGTCTATTCCGACGACCTCGTGTGTATGCGGAATCTCCACCATCGCAACTACCACCTG	1584
QY	514	-----AlaAlaIle-----	516
Db	1584	GTGGTTCGAACATCTGAATTGTATGCTGGGTTCACAGCGCCAAAGACAGACGCACT	1643
QY	516	eThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnArgGlnMe	536
Db	1644	AACGGACTCCCATCATTTCTAGCTGTAGGATAATA-----AATGACGCCGACCT	1694
QY	536	tLeuPro 538	
Db	1695	AGTCCCC 1701	
RESULT 5			
US-09-981-282-4			
; Sequence 4, Application US/09981282			
; Patent No. 6641819			
; GENERAL INFORMATION:			
; APPLICANT: Mengeling, William L			
; APPLICANT: Vornwald, Ann			
; APPLICANT: Lager, Kelly			
; APPLICANT: Roof, Mike			
; APPLICANT: Burkhardt, Kelly			
; APPLICANT: Gorceyca, David			
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE			
; FILE REFERENCE: 27093B			
; CURRENT FILING DATE: 2002-08-19			
; CURRENT APPLICATION NUMBER: US/09/981,282			

```
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 15413
; TYPE: RNA
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-09-981-282-4

Alignment Scores:
Pred. No.: 0.124 Length: 15413
Score: 111.00 Matches: 86
Percent Similarity: 32.92% Conservative: 47
Best Local Similarity: 21.29% Mismatches: 137
Query Match: 3.26% Indels: 134
DB: 4 Gaps: 21

US-09-932-678-2 (1-651) x US-09-981-282-4 (1-15413)

QY 189 AlaLeuGlnIleAlaArgTyrValProSerThrProTrpPheLeuMetPro----- 206
Db 8411 GCAUUAAGAGACCCUCUCCAAAGUAGACUUGUCCACCAAGGCUUUGUUUACUGAGUU 8470

QY 207 ---IleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys--- 224
Db 8471 CUUCCCUUGCGUAGUAGUACUGUUGUCUAGUGGGU-----AAGUGCCG 8518

QY 225 TyrValHisAsnLeuArgIleSerValTyr----- 235
Db 8519 CCGGUUUAU-----CGGCUUCCACUACUCCUCCCAAGAUUUCUAGGCGUAAUA 8569

QY 236 -----PheProThrLeuArgHisGluIleLeuGluLeuIleLeuGluLysLeu 251
Db 8570 AAUGGAACAGGUUCCACCAAGGAC-----AUCCAGAGCGUC 8608

QY 252 LeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluThrAlaThr 271
Db 8609 CUGAAUUGCGUUGUGGCGCA---CAGGCGGUGCGGAAACUGGCAACUGUUAAC 8665

QY 272 -----GlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsn 285
Db 8666 CUUUGACCCUACAGAAACAGUUGUGG----- 8695

QY 286 MetAspGluAspGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGln 305
Db 8696 -----AAGAAGAGAGUAGGACAAUACUGGCAACAAUUAUUAUUAUUAUUAUUA 8740

QY 306 MetValHisProValaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMet 325
Db 8741 CUGGCCCCCGGCGAGC-----UUGAGUGUGUGUACCCAGCGGCUUUAUG 8785

QY 326 LysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArg 345
Db 8786 AAAAAGGCG-----UUUAACUGGCGCAUUGCCCGUAAACAA 8827

QY 346 AspLeuIleAsnIlePheAspLysLeuLeuProThrHisAlaSerCysHisValGln 365
Db 8828 -----UUUAAAGAGUUCAGACUCCGUGUAGGCGAGGCG----- 8863

QY 366 PhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHis 385
Db 8864 -----CUUGAAGCUGUACUUGCAUCC 8884

QY 386 LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsn 405
Db 8885 UGC-----GAUGCUCCACACUCCGAAUUGUCCGUGUUGCGGCGGCGGCGGCGG 8929

QY 406 TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys 425
Db 8930 CUUCUUUAUGAACUUGCGUGUGAGAGACCGUGCGUGUACGUGUAGUUAACUGCGUC 8989

QY 426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLys 445
Db 8990 CAGACCUACUGGUC-----ACGAGUCCGCGGCGAGUAAACU 9025

; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 15413
; TYPE: RNA
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-09-981-282-2

Alignment Scores:
Pred. No.: 0.125 Length: 15424
Score: 111.00 Matches: 86
Percent Similarity: 32.92% Conservative: 47
Best Local Similarity: 21.29% Mismatches: 137
Query Match: 3.26% Indels: 134
DB: 4 Gaps: 21

US-09-932-678-2 (1-651) x US-09-981-282-2 (1-15424)

QY 189 AlaLeuGlnIleAlaArgTyrValProSerThrProTrpPheLeuMetPro----- 206
Db 8412 GCATTAAGAGACCTCTCCAGTATGACTTGTCACCAAGGCTTTGTTTACCTGGAGTT 8471

QY 207 ---IleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys--- 224
Db 8472 CTTGCGCTTGCGTAGTACTCTTGTCTCATGTGGGT-----AAGTGCAG 8519

QY 225 TyrValHisAsnLeuLeuArgIleSerValTyr----- 235
Db 9025 -----AAGTGCAG 8519
```





```
Db 8787 AAAAAAGGCG-----TTTAACTCGCCCAATGGCTCGGTAAACAA----- 8828
QY 346 AspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGln 365
Db 8829 -----TTTAAAGAGCTTCAGACTCCGGTCTTAGCGCAGGTGC----- 8864
QY 366 PhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHis 385
Db 8865 -----CTTGAAGCTGATCTTGATCC 8885
QY 386 LeuTrpLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsn 405
Db 8886 TGC-----GATCGCTCCACACCTGCAATTCGCTGTTGCGCAAT 8930
QY 406 TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys 425
Db 8931 CTTCTTATGAATCGCTGCTGAGAGACCTGCGCTGCTGATCGTGTGTGAAGTCTGCTGC 8990
QY 426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLys 445
Db 8991 CACGACCTACTGGTC-----ACGAGTCCGGCGCAGTAAC 9026
QY 446 AlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyr 465
Db 9027 AAGAGAGTGGCTGTCTGCTGCGCAGCGATCACTTCTGTGTCCAACACCATTTACAGC 9086
QY 466 ThrPheValPheArgHisLysGlnLeuLeuSer-----GlyAsnLeuLysGluGly 482
Db 9087 TTGGTGATATGTCACACACATGGTGCTCAGTTACTTTAAAGTGGTCACCTCATGGC 9146
QY 483 LeuGlnTyrLeuGln---SerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPro 501
Db 9147 CTTCTGTTCTACAGACGACTGAAGTTTGAGGAC-----ATGCTCAAGGTTCAACCC 9200
QY 502 LeuLysIleCys-----LeuProSerValValAsnPhe 512
Db 9201 CTGATGCTATATCGGACGACCTCGTACTGTATGTCGAGTCTCCACCATGCCAACTAC 9260
QY 513 Phe-----Ala 514
Db 9261 CACTGGTGGTGAACATCTGAACCTGATGCTGGGTTTCAGACGACCCAAAGAGACA 9320
QY 515 AlalleThrAsnLysTyrGlnLeuValPheCysTyrThrIleGluArgAsnAsnArg 534
Db 9321 GCCATAACAGACTCGCCATCTTCTAGGCTGTAGGATAATA-----AATGACGC 9371
QY 535 GlnMetLeuPro 538
Db 9372 CAGCTCGTCCCT 9383
```

## RESULT 8

```
US-09-470-661A-1
; Sequence 1, Application US/09470661A
; Patent No. 650662
; GENERAL INFORMATION:
; APPLICANT: Pfizer Products Inc.
; TITLE OF INVENTION: AN INFECTIOUS CDNA CLONE OF NORTH AMERICAN PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME (PRRS) VIRUS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: PCI0278A
; CURRENT APPLICATION NUMBER: US/09/470,661A
; CURRENT FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 15450
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: cdna
; OTHER INFORMATION: corresponding to No. 6500662th American Porcine
; OTHER INFORMATION: Respirative And Respiratory Syndrome (PRRS) Virus
; OTHER INFORMATION: Genome.
```

```
US-09-470-661A-1
Alignment Scores:
Pred. No.: 0.125 Length: 15450
Score: 111.00 Matches: 128
Percent Similarity: 32.34% Conservative: 67
Best Local Similarity: 21.23% Mismatches: 231
Query Match: 3.26% Indels: 178
DB: 4 Gaps: 27

US-09-932-678-2 (1-651) x US-09-470-661A-1 (1-15450)
QY 2 AlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAlaVal 21
Db 7892 GCAGTTCCTTCCTTATAGATGTCGTGATCTCGGTGCTGACGCACTCTCTAAGTTACTTC 7951
QY 22 LysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPhe 41
Db 7952 GCTCGTACGCGCGCGGGAACACTGGGATCGATGGCAGCTTTGG-----GACTTT 8002
QY 42 PheAsnSerProProArgLysThrValArgPheGly----- 53
Db 8003 GAGCGCGAGGCCACCAAGAGGAAATTGCCTCAGTGCACAATAATACAGGCTTTGAC 8062
QY 54 -----GlyThrValThrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAsp 70
Db 8063 ATTAGCGCGCGCGACGACCTGAAATGGTCTCCCTTACAAG----- 8104
QY 71 PheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleLeuAsn 90
Db 8105 -----CTGTACCTGTGTAGGGGCAACCCCTGAGCGGGTAAAA 8140
QY 91 TrpLeuLeuGlu-PheArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeu 110
Db 8141 GGAGTTTACAGAAACAGAGTTTGGAGACATACCTTACAAACCCCCAGTGCACATGGA 8200
QY 110 eSerIleLeuLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLe 130
Db 8201 ASCCCAGTGCACGCGCTGCCTGC-CTCAGCGCCCAATGCCACTCCGGTACTGAT----- 8254
QY 130 uAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMe 150
Db 8255 -----GGCGGCTGTCTTGTGCTACTACCATG-----CCCTCGCGTTTGA 8295
QY 150 tileAlaSerHisPheValProArgValIleIleLysGluGlyAspValAspValse 170
Db 8296 ATTGTATGTACGACCATTCACGCTGTCTCCT-----GATTATCT 8337
QY 170 rAspSerAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLe 190
Db 8338 TGACTCTAGGCTGACTGCCCCAAACAGTTGACAGACGCGCTGTGAGATGCCGATT 8397
QY 190 uGlnIleAlaArgTyrValProSerThrProTrpPheLeuMetPro-----Il 207
Db 8398 GAGAGACCTCTCCAAGTATGACTTGTCCACCAAGGCTTTGTTTACCTGGGTTCTTCG 8457
QY 207 eLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys---TyrVa 226
Db 8458 CCTTGTGCGTAAGTACCTGTTTGGCCATGTGGGT-----AAGTCCCGCCCT 8505
QY 226 IhisAsnLeuLeuArgIleSerValTyr----- 235
Db 8506 TCAT-----CGGCTTCCACTTACCTGCCAAGAATTCTATGGCTGGAATAAATGG 8556
QY 236 -----PheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLeu 253
Db 8557 GAACAGGTTTCCAAACCAAGGAC-----ATTACAGCGTCCCTGA 8595
QY 253 sleuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThr----- 271
Db 8596 AATCAGCGTCTGTGCGCA--CAGCGCGTGCAGAAACTGGCAAACTGTACCCCTTG 8652
QY 272 -----GlnThrCysGlyThrAspSerThrGluGlyLeuPheAsnMetAs 287
```

```
Db 8653 TACCTCAGAAACAGTATTGTGGG----- 8677
Qy 287 pGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetVa 307
Db 8678 -----AAGAAAGAACTAGGACAATCTCGGCACCAATAATTTCATTGGTGGC 8727
Qy 307 lHisProValAlaGluArgLeuAspLleLeuMetSerLeuValLeuSerTyrMetLysAs 327
Db 8728 CCACCGGGCAGCG-----TTGAGTGGTGTCCACCAGGGCTTCATGAATAA 8772
Qy 327 pValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLe 347
Db 8773 GGCG-----TTTAACTCGCGCATCGCCCTCGGGAATAACAA----- 8809
Qy 347 uIleAsnIlePheAspLysLeuLeuProThrHisAlaSerCysHisValGlnPhePh 367
Db 8910 -----TTTAAGAGCTACAGACTCCGCTCTTAGGACGGTGC----- 8845
Qy 367 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuHisLeuTr 387
Db 8846 -----CTTGAAGCTGATCTTGATCTCTGT-- 8869
Qy 387 pLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIl 407
Db 8870 -----GATCGATCCACACTGCAATTGTCGCTGTTGCGCCATCTTCT 8916
Qy 407 eGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerLeuAs 427
Db 8917 TTATGAATCCCTGCTGTAAGAGCCTACCTACCTGCTGTAACCTGCTGTAACCTGCTGTA 8976
Qy 427 pLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPh 447
Db 8977 CCTATTGTC-----ACGCACTCGCGCGAGTACTAAGAG 9012
Qy 447 eCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh 467
Db 9013 GGGTGGCTATCGTCTGCGCAGCCGATCACTTCTGTCTTAACACCATTTACAGCTTGGT 9072
Qy 467 eValPheArgHisLysGlnLeuLeuSer-----GlyAsnLeuLysGluGlyLeuGl 484
Db 9073 GATATATGCAAGCAGCATGGTGTCTAGTACTTTAAAGTGTGTCACTCCCTCATGCGCTTCT 9132
Qy 484 nTyrLeuGln---SerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLy 503
Db 9133 GTTCTTACAGACCACTGTAAGTTCGAGC-----ATGCTCAAAGTCCAAACCTGTAT 9186
Qy 503 sIleCys-----LeuProSerValValAsnPhePhe-- 513
Db 9187 CGTCTATTGCGAGCAGCTCGTGTGTATGCGGAATCTCCACCATCGCGAACTACCACTG 9246
Qy 514 -----AlaAlaIl 516
Db 9247 GTGGTGCAGCATCTGAATTGATGCTGGTTCAGACGACCAACAAAGACAGCAT 9306
Qy 516 eThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMe 536
Db 9307 AACGACTCGGCATCTTCTTAGCTGTAGGATAATA-----AATGACGCGCAGCT 9357
Qy 536 tLeuPro 538
Db 9358 AGTCCCC 9364

RESULT 9
US-09-981-282-3
; Sequence 3, Application US/09981282
; Patent No. 6641819
; GENERAL INFORMATION:
; APPLICANT: Mengeling, William L
; APPLICANT: Vorwald, Ann
; APPLICANT: Lager, Kelly
; APPLICANT: Roof, Mike
; APPLICANT: Burkhardt, Kelly
```

```
; APPLICANT: Gorcyca, David
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE
; TITLE OF INVENTION: BASED ON ISOLATE JA-142
; FILE REFERENCE: 27093B
; CURRENT APPLICATION NUMBER: US/09/981,282
; CURRENT FILING DATE: 2002-08-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 15413
; TYPE: RNA
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-09-981-282-3

Alignment Scores:
Pred. No.: 0.524 Length: 15413
Score: 106.00 Matches: 127
Percent Similarity: 32.17% Conservative: 67
Best Local Similarity: 21.06% Mismatches: 232
Query Match: 3.11% Indels: 178
DB: 4 Gaps: 27

US-09-932-678-2 (1-651) x US-09-981-282-3 (1-15413)
Qy 2 AlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAlaVal 21
Db 7910 GCATUUCUUGCGUUAUAGACGUCUUAUUCUCCGCGCUGAUGCAUCCCAAGUACUC 7969
Qy 22 LysLysLeuGlyAlaSerArgThrGlyLysSerAsnMetArgAlaLeuGluAsnAspPhe 41
Db 7970 GCCCGCACGGCGGGGAAACACUGGAUGCAUGCGCAGCGCUUUGG-----GAUUU 8020
Qy 42 PheAsnSerProProArgLysThrValArgPheGly----- 53
Db 8021 GAGCCCGAGCCCAAAAGAGAAUUGCACUGCGCGCAAAUAAUACAGCGCUUGGAC 8080
Qy 54 -----GlyThrValThrGluValLeuLeuLysTyrLysLysGlyThrAsnAsp 70
Db 8081 AUUAGCGCGCGCGACGCGACUAAUUGUUCUUAUAG----- 8122
Qy 71 PheGluLeuLeuLysAsnGlnLeuLeuAspProAspLysAspAspGlnIleLeuAsn 90
Db 8123 -----CUGUACCCUGCGAGGCGCAACCCUGAGCGGGAUAA 8158
Qy 91 TrpLeuLeuGlu-PheArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIl 110
Db 8159 GGAGUUUACAGAAUACAGGUUGGAGAUUACCUAAUAAACCCCGAGUGACACUGGA 8218
Qy 110 eSerIleLeuLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLe 130
Db 8219 AGCCGAGUGCAGCGCGGCGUGCGC- CUCACGCCCAUAGCCACGCGGUGACUGAU- 8272
Qy 130 uAlaPheLeuGlyAsnLeuValSerAlaGlnThrValLeuLeuArgProCysLeuSerMe 150
Db 8273 -----GGGCGCCUGCGUGCGCACACUAG-----CCUCCGGUUUUGA 8313
Qy 150 tIleAlaSerHisPheValProArgValIleIleLysGluGlyAspValAspValSe 170
Db 8314 GUUGUAGUACCGACCAUUCACGCGUGUCUUCU-----GAUUAUCU 8355
Qy 170 rAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLe 190
Db 8356 UGAUUCUAGCGCUGACUGCCCAACAGUUGACAGACGCGUGUGAGGACGCGCAU 8415
Qy 190 uGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetPro-----Il 207
Db 8416 AAGAGACCUCCACAGAUAGACUUGUCCACCAAGGCUUUUUUUUACUGAGUUCUUG 8475
Qy 207 eLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys---TyrVa 226
Db 8476 CCUUGUGCGUAAGUACCUUGUUGCUCAUGUGGU-----AAGUGCCCGCGCU 8523
Qy 226 lHisAsnLeuLeuArgIleSerValTyr----- 235
```



```

      |||||
8357 TGATTCTAGGCTGACTGCCCCCAACAGTTGACAGACACGGCTGTGTAGGACGGCGCAT 8416
      |||||
190 uGlnIleAlaArgTyrValProSerThrProTyrPheLeuMetPro-----II 207
      |||||
8417 AAGAGACCTCTCCAAGTATGACTGTCTCACCCAGGCTTTGTATTACCTGGAGTTCCTCG 8476
      |||||
207 eLeuValGlnLysPheProPheValArgLysSerGluArgThrLeuGluCys--TyrVa 226
      |||||
8477 CTTGTGTGCTAAGTACCTGTTGTCTCATGTGGT-----AAGTCCCGCCCGT 8524
      |||||
226 lHisAsnLeuLeuArgLysSerValTyr----- 235
      |||||
8525 TCAT-----CGGCTTCCACTTACCTCCGCCAAGAAATCTATGGCTGGAAATAATGG 8575
      |||||
236 -----PheProThrLeuArgHisGluIleLeuGluLeuIleLeuGluLysLeuLeuL 253
      |||||
8576 GAACAGGTTTCCAAACCAAGGAC-----ATCCAGAGCGTCCCTGA 8614
      |||||
253 sLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThr-- 271
      |||||
8615 AATCGACGTTCTGTGCGCA--CAGGCGTTTCGGGAAACTGGCAAACTGTACCCCTTG 8671
      |||||
272 -----GlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAs 287
      |||||
8672 TACCTCAAGAAACAGTATTGTGG----- 8696
      |||||
287 pGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetVa 307
      |||||
8697 -----AAGAAGAAGACTAGACAAATACTCGGCACCAATAACTTTCATTCGCTGGC 8746
      |||||
307 lHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAs 327
      |||||
8747 TCACCGGCGACG-----TTGAGTGGTGTCCACCAGGCGCTTCATGAATAA 8791
      |||||
327 pValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLe 347
      |||||
8792 GGCG-----TTTAACCTGCCATTCGCTCGGTAACAAACAA-- 8828
      |||||
347 uIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePh 367
      |||||
8829 -----TTTAAGAGCTTCAGACTCCGCTCTAGGCAGGTGC----- 8864
      |||||
367 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTr 387
      |||||
8865 -----CTTGAAGCTGATCTTGCACTCTGCATCCTGCG-- 8888
      |||||
387 pLysLysLeuGlnAspProSerAsnProAlaIleLeuArgGlnAlaAlaGlyAsnTyrIl 407
      |||||
8889 -----GATCGCTCCACCTGCATTTGTCGCTGGTTCGCGCCCACTTCT 8935
      |||||
407 eGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAs 427
      |||||
8936 TTATGAACCTTCCTGTGTGTAAGACACACGACCGCTGTGTAAGTCTGCTGCCAGA 8995
      |||||
427 pLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPh 447
      |||||
8996 CCTACTGGTC-----ACGCAAGTCCGCGCAGTAACTAAGAG 9031
      |||||
447 eCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh 467
      |||||
9032 AGTGGCTGTGCTGTCGCGACCGATGATCACTTGTGTCCAAACACCATTTACAGCTGGT 9091
      |||||
467 eValPheArgHisLysGlnLeuLeuSer-----GlyAsnLeuLysGluGlyLeuGl 484
      |||||
9092 GATATATGCACACACATGGTGTCTCAGTTACTTTAAAAGTGTCTCACCTCATGCGCTTCT 9151
      |||||
484 nTyrLeuGln---SerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuL 503
      |||||
9152 GTTCTTCAAGACACGAGTGAAGTTTGAGGAC-----ATGCTCAAGGTTCAACCCCTGAT 9205
      |||||
503 sIleCys-----LeuProSerValValAsnPhePhe-- 513
      |||||

```

Db 9206 CGTCTATTGGAGACGCTCGTACTGTATGCGAGTCTCCACCATGCCAAAATCACTACCTG 9265  
 Qy 514 -----AlaAlaI 516  
 Db 9266 GTGGGTTGAACATCTGAACCTGATGCTGGGTTTTCAGACGACCCCAAGAGACAGCCAT 9325  
 Qy 516 eThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMe 536  
 Db 9326 AACAGACTGGCCATCATTTCTAGGCTGTAGGATAATA-----AATGGACGCCAGCT 9376  
 Qy 536 tLeuPro 538  
 Db 9377 GGTCCCT 9383  
 RESULT 11  
 US-10-143-186-1  
 ; Sequence 1, Application US/10143186  
 ; Patent No. 6660513  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MENGELING, WILLIAMS L.  
 ; APPLICANT: VORWALD, ANN  
 ; APPLICANT: LAGER, KELLY  
 ; APPLICANT: ROOF, MIKE  
 ; APPLICANT: BURKHART, KELLY  
 ; APPLICANT: GORCYCA, DAVID E  
 ; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE,  
 ; FILE REFERENCE: 27093a  
 ; CURRENT APPLICATION NUMBER: US/10/143,186  
 ; PRIOR APPLICATION NUMBER: US/09/461,879  
 ; PRIOR FILING DATE: 1999-12-15  
 ; PRIOR APPLICATION NUMBER: 09/298,110  
 ; PRIOR FILING DATE: 1999-04-22  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 15424  
 ; TYPE: DNA  
 ; ORGANISM: Porcine reproductive and respiratory syndrome virus  
 US-10-143-186-1  
 Alignment Scores:  
 Pred. No.: 0.525 Length: 15424  
 Score: 106.00 Matches: 127  
 Percent Similarity: 32.17% Conservative: 67  
 Best Local Similarity: 21.06% Mismatches: 232  
 Query Match: 3.11% Indels: 178  
 Db: 4 Gaps: 27  
 US-09-932-678-2 (1-651) x US-10-143-186-1 (1-15424)  
 Qy 2 AlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaSerSerAlaVal 21  
 Db 7911 GCATTCCTTCGCTTATAGACGCTCTTAATCTCCGCGCTGATCATCTCCCAAGTTACTC 7970  
 Qy 22 LysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPhe 41  
 Db 7971 GCCGCGCACGGCGCGGAAACACTGGGATCGATGCGCAGCTTGG-----GATTTT 8021  
 Qy 42 PheAsnSerProProArgLysThrValArgPheGly----- 53  
 Db 8022 GAGCGCGAGGCCACTAAAGAGAAATTCAGTCCGCAAAATATACAGGCTTGTGAC 8081  
 Qy 54 -----GlyThrValThrGluValLeuLysTyrLysGlyGluThrAsnAsp 70  
 Db 8082 ATTAGCGCGCGGACGCGACCTGAAATTTGCTTCCTTATAAG----- 8123  
 Qy 71 PheGluLeuLysAsnGlnLeuLeuAspProAspIleLysAspGlnIleAsn 90  
 Db 8124 -----CTGTACCCTCTCAGGCGCAACCTCTGAGCGGTAATA 8159  
 Qy 91 TrpLeuLeuGlu-PheArgSerIleMetTyrLeuThrLysAspPheGluLeuI 110



```
Db 501643 GAATGTGATTAATAACATTACCATTAATAAATAAGGATTGTACCAATTTTCCT----- 501593
Qy 155 PheValProProArgValIleIleLysGluGlyAspValAspValSerAspSerAsp 174
Db 501593 ----- 501593
Qy 175 GluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleAla 194
Db 501592 -----AATAACTTA-----AATTATATGTCGGAAGGAATATGATGAAGTTCAA 501545
Qy 195 ArgTyrValProSerThrProTyrPheLeuMetProIleLeuValGluLysPheProPhe 214
Db 501544 GAGCGGATAGTTGACTT-----TTTCAAGCA 501518
Qy 215 ValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeu-----LeuArgIle 232
Db 501517 GTTAGTTCCAAAGAAAGAACGATTAAATTTAGTGGTTGAAATATTTGATGATTTTCGTAAT 501458
Qy 233 SerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleGluLysLeuLeu 252
Db 501457 CAATTCACCAATATGAAAGAAAGAAATGAGCTTTTAAACCAACAATTT---AAAGCCATT 501401
Qy 253 LysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGln 272
Db 501400 GAATTA---GTTGCCAAATGAACAGATGTGATGATTTAATTTATGAGCTTAATGAG 501344
Qy 273 ThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThr 292
Db 501343 AACTGCGATCTAATT-----ACTAAATAATTAATACTTTAAACCATCAAGTTTAGCT 501290
Qy 293 GluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGlu 312
Db 501289 CAGAATCAAAATCAAT-----TATGAA 501269
Qy 313 ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAsp 332
Db 501268 AAACGTGTTTACTGATGAGTGAACAAACAAATCAAGCTTTAGAA----- 501227
Qy 333 GlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAsp 352
Db 501226 ---AAACTGAAATAATCACTAACT--GATTTACTTAATTTAGCAGTTTAATAAGTTTAT 501173
Qy 353 LysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCys 372
Db 501172 -----GACTATCGCTTTTGTGCTATACCTTTAGAT 501143
Qy 373 SerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTyrLysLysLeuGlaAsp 392
Db 501142 GATTTCCGATTAAATTTTTCAG-----ATTGAAATTTTAATTAGCGAT---CAAGAT 501092
Qy 393 ProSerAsnProAla----- 397
Db 501091 TTAAGGTTCTCTGTAAACATTCAGATAATTAATAATCGAAATTAGAATTTGATAAC 501032
Qy 398 IleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIle 417
Db 501031 ATCTTCACTGAAGCTAATAATACTAATTTAGTACTTTTAAATTTGTTTAGTGAAAGATA 500972
Qy 418 ProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTyrLeuHisIleTyrLeu 437
Db 500971 GATTTATGCAAAATTCATTC----- 500951
Qy 438 AsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyr 457
Db 500950 -----ACTGAAGTTATTTGTGATGTTGTTAGTTAGTTAACTTAAGG 500909
Qy 458 SerAlaCys---GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSer 476
Db 500908 CTAATGTGCAACCAAGCTTTTCTTTTGTGCAAAATAAATATATCGTCAAAAG-----AGT 500858
Qy 477 GlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeu 489
Db 500858 -----
```

```
Db 500857 AGTCATATCGATAGATCATTTAGAACAGTTAAACAAACTTA 500819
RESULT 13
US-08-329-681A-3
; Sequence 3, Application US/08329681A
; Patent No. 5650294
; GENERAL INFORMATION:
; APPLICANT: Kurth, Roland
; APPLICANT: Philippsen, Peter
; APPLICANT: Steiner, Sabine
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: No. 5650294el Promoter Region
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.4 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/329,681A
; FILING DATE: 26-OCT-1994
; CLASSIFICATION: 435
; CLASSIFICATION: C12P 25/00
; CLASSIFICATION: C12N 1/14
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2115 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-329-681A-3
Alignment Scores:
Pred. No.: 0.0356 Length: 2115
Score: 103.50 Matches: 50
Percent Similarity: 40.54% Conservative: 25
Best Local Similarity: 27.03% Mismatches: 71
Query Match: 3.04% Indels: 39
DB: 1 Gaps: 10
US-09-932-678-2 (1-651) x US-08-329-681A-3 (1-2115)
Qy 23 LysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsn----- 39
Db 880 AAAATGGTGAAGAAAGAAAGACTCAGTTTCGAGGCCGGGATTAATTCACATGGATGCT 939
Qy 40 AspPhePheAsnSerPro---ProArgLysThrVal---ArgPheGlyThrVal--- 56
Db 940 GATTTATATGGGTATAAAATGGCTCGCGATAATTCGGGCAATCAGGTGCGACATCTAT 999
Qy 57 -----ThrGluValLeuLeuLysTyrLysLysGlyGluThr 68
Db 1000 CGATTGTATGGGAAGCCCGATGGCCAGATGTTTCTGAACACATGCAAGGTAGCGTT 1059
Qy 69 AsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleLysAspGlnIle 88
Db 1060 GCCAAT-----GATGTTACAGATGAGATGTCG 1086
Qy 89 ---IleAsnTrpLeuLeuGluPheArg-----SerSerIleMetTyrLeuThrLysAsp 105
Db 1087 AGACATAAATCGCTGACGGAATTTATGCTCTTCCGACCATCAAGCATTTTATCCGTA 1146
Qy 106 PheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrVal 125
Db 1147 CCTGATGATGATGTTTACTCAACCATCGATCCCGGGAAAAACAGCAATTCACAG---GTA 1203
```





Db 4901 CGAATACAAAGAGAGTTGATTTCTGCTGCTGGTGTCTGCTGTT----- 4943  
QY 392 pProSerAsnProAlaIleIleArgGlnAlaLaGlyAsnTyr----- 406  
Db 4944 -----ACATTTGTTTCTGCTCATCAAAATTTGTCATATCCFAAATATTATACITTT 4993  
QY 407 -IleGlySerPheLeuAlaArgAlaLysPheIle-----ProLeuIleThrVa 422  
Db 4994 TACTGCTGCTGCTATGAAGAAACAAAGGTTATTAGACACATTTGAACCCAGCTTTGGCTTTT 5053  
QY 422 llySerCysLeuAspLeuValAsnTrp----- 432  
Db 5054 CCAATTGGAAATTTGGAGAGATTACCCCAATTTGCATATCAACCAATTTTCCATTAACAACAG 5113  
QY 433 -----LeuHisIleTyr-----LeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAs 449  
Db 5114 AAACATCCATGATATATGATGCAATTTGGGAAGAAATGCTCTCTCGATAAAGATTT----- 5168  
QY 449 pValAlaLeuHisGlyPropheTyrSerAlaCysGlnAlaValPheTyrThrPheValph 469  
Db 5169 -----TT 5170  
QY 469 eArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLe 489  
Db 5171 CACCAGAGGGATTATTAGAACCGGTGTTCTTAAAGAGACATTTAGCATTAGTGAATATTT 5230  
QY 489 uAsnPheGlu-----ArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuPr 507  
Db 5231 GATTGCTGAATCCAACAGATTAATGAATCATATTTTGGATACTTTTAGAAGTTATTGACAC 5290  
QY 507 oSer-----ValValAsnPhe-----PheAlaAlaIleTh 517  
Db 5291 TTTCAATTCGATTAAACCATATTTTCAATTAACITTTTCCAAATGCTTTCAATGTTCAAGC 5350  
QY 517 rAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLe 537  
Db 5351 TTCAGATGTTGAGGCTGCTTT---GGATCATTTTAGAAGATTTTGGTAGAAGATTATG 5407  
QY 537 uProValIleArgSerThrAlaGlyAspSerValGlnIleCysThrAsnProLeuAs 557  
Db 5408 GAGATTA-----ACAGTTACTGCTGCTGCTGAATTAGAATTTGCTGCTACTGCTCCTCAAG 5461  
QY 557 pThrPhePhePro 561  
Db 5462 TACTGCTGTTCCCA 5474

## RESULT 15

US-09-134-001C-335  
; Sequence 335, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 335  
; LENGTH: 1692  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-335

## Alignment Scores:

Pred. No.: 0.0323 Length: 1692  
Score: 102.50 Matches: 104  
Percent Similarity: 35.32% Conservative: 74  
Best Local Similarity: 20.63% Mismatches: 147

Query Match: 3.01% Indels: 179  
DB: 4 Gaps: 25  
US-09-932-678-2 (1-651) x US-09-134-001C-335 (1-1692)  
QY 57 ThrGluValLeuLeuLysTyrLysGlyGluThrAsuAspPheGluLeuLeuLysAsn 76  
Db 415 ACGCAATCTTTACTTAAGCAAAA-----TATCATCTT----- 447  
QY 77 GlnLeuLeuAspProAspIleLysAspGlnIleLeuAsnTrpLeuGluPheArg 96  
Db 448 CAACTATTAGAT---GATTATGACACAATCAGTATTACAGATTACTTAAATCAATATCAA 504  
QY 97 SerSerIleMetTyrLeu-----ThrLysAspPheGluGlnLeuIleSer----- 111  
Db 505 CTTTCTTATTAACCAATATAAAATAAACGTAAAGAAATTAGAGAAATTAGAAATCCCGGAC 564  
QY 112 -----IleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyr 129  
Db 565 CAAGCTTTATTACACGATTA-----GACTTA 591  
QY 130 LeuAlaPhe-----LeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeu 148  
Db 592 ATGAAATTTCAATTAGAGAACTAACCGAAGCTTCA----- 627  
QY 149 SerMetIleAlaSerHisPheValProArgValIleIleLysGluGlyAspValAsp 168  
Db 628 -----CTGAAAGAAAGCGCAAGTGAC 648  
QY 169 ValSerAspSerAsp-----AspGluAspAspAsnLeuProAlaAsnPheAsp 184  
Db 649 CAACTTGAATCCGATATTAAAGAAATTCAAAACCTCGAAAAATTAATCTAGCTTTAAAC 708  
QY 185 ThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeu 204  
Db 709 AATGCACATCAAGTTCTA-----ACTGATGAAAGTGCA 741  
QY 205 MetProIleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCys 224  
Db 742 ATACCC----- 747  
QY 225 TyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeu 244  
Db 748 -----GATAGTTGTGACGAATTAAGCAACTACTTGTCAACAGATTAAT----- 789  
QY 245 GluLeuIleLeuLysLeuLysLeu-----AspValAsnAlaSerArgGlnGly 262  
Db 790 GATATCGTTCAGAAAAAATTCGTAAAGATTAAAGAGGACATTAATCAATTTTACTATTG 849  
QY 263 IleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGly 282  
Db 850 CTAGAAGATGCAAGCATGAATTT----- 873  
QY 283 LeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArg 302  
Db 874 -----TACGACGAATGGCTTAACACTGAATTCGAT----- 903  
QY 303 LeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu 322  
Db 904 ---GAGCAAGTTTTAAATAGTAGTAGAATCCAGAATGAATTTACTTAATAATTTAAAAAGT 960  
QY 323 SerTyrMetLysAspValCys-----TyrVal 331  
Db 961 AAATATGTTAAGGATTTACTGAATTTATGCTTATCAGAGTAACCTTGCAATTTGAAATTT 1020  
QY 332 AspGlyLysValAspAsn-----GlyLysThrLysAsp 342  
Db 1021 GAT---AAAATAGAAAACATATGAACAAAGTACATCACAATTAAGGGAGAAAATTAACG 1077  
QY 343 LeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeu----- 355  
Db 1078 CTTTATAACAGAGTAGATATATAGGAAAAAAACTTTCTCAAGAACGTAGGCGGTAGCG 1137

```
QY 356 -----LeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSer 373
Db 1138 AGAGAGTTAAGGACCATATTTCTCTGAAATACAA----- 1173
QY 374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393
Db 1174 ---AATTACAAATGAAGATGCTAACCTTGAAATTCGTTTAAACCATTAGATGAACCT 1230
QY 394 Ser-----AsnPro----- 396
Db 1231 ACAATTGAAGTATTGAATTTCTGGAAATTTTAATTAGTCCAAATCGTGGTGAACCACTT 1290
QY 397 ---AlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLys 415
Db 1291 AAAAGTCITTAATAAATCGCTTCAGCGGTGAACTTCAAGAATTATGTAGTCTTAAAA 1350
QY 416 PheIleProIleIleThrValLysSerCysLeuAspLeu-LeuValAsnTrpLeuHisI 435
Db 1351 AGTATATTGTTAATCACGCGGCCAACCGCGATTCTTTTGATGAAGTTGACTCGGCT 1410
QY 435 eTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyPr 455
Db 1411 GTATCTGGTCAAGCAGCACAAAAATGGCTGAAAAAATCGAGATATTGCTCAAT----- 1465
QY 455 oPheTyrSerAlaCysGlnAlaValPheTyrThrPheVal-PheArgHisLysGlnLeuL 475
Db 1466 -ATATACAAAGTCA-----TTTGATTTTACACTTACCTCAGGTAGTCTTCAATGAGTGAC 1518
QY 475 euSerGlyAsnLeuLysGluGlyLeuGln-----TyrLeuGlnSerLeuAsnP 491
Db 1519 CATCATCTTCTAATAAGCAAGGCATCCAAATGCCGATAGAACTACAACCTCAAGTCAAAGAA 1578
QY 491 heGluArg 493
Db 1579 TTGAAAGA 1586
```

Search completed: July 18, 2004, 12:13:50  
Job time : 411 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 06:23:07 ; Search time 670 Seconds  
(without alignments)  
4127.727 Million cell updates/sec

Title: US-09-932-678-2

Perfect score: 3403  
Sequence: 1 MAAPLLHTRLPDAAASSA.....RSPSSVSGPPVLVYMQPSPL 651

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgm2.1/UsPTO.spool/US09932678/runat.15072004.103603.5196/app.query.fasta\_1.839  
-DB=N Geneseq 29Jan04 -QWMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09932678 @CGN 1.1 470 @runat.15072004.103603.5196 -NCPU=6 -ICPU=3  
-NO MWAP -LARGQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04:\*  
1: geneseq1980s:\*  
2: geneseq1990s:\*  
3: geneseq2000s:\*  
4: geneseq2001as:\*  
5: geneseq2001bs:\*  
6: geneseq2002s:\*  
7: geneseq2003as:\*  
8: geneseq2003bs:\*  
9: geneseq2003cs:\*  
10: geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3403	100.0	2068	6 ABV72159	Nucleotid
2	3391	99.6	2040	3 AA98384	Human RNA
3	2839	83.4	1770	6 ABS78724	DNA encod
4	2162	63.5	1418	3 AAC76937	Human ORF
5	2001	58.8	1582	9 ADD29697	Human tum
6	1973	58.0	1423	4 AAS01562	Human sec
7	1763	51.8	1498	4 AAS44584	Human ful
8	1313	38.6	3169	5 AAS87641	DNA encod

9	1280.5	37.6	1989	6	ABK47950	Human tra
10	1205	35.4	3175	5	AA92254	DNA encod
11	1181	34.7	1017	5	AA87636	DNA encod
12	1124	33.0	1461	5	ABA83038	Human tra
13	1074	31.6	2410	5	AA92255	DNA encod
14	1000	29.4	2493	6	AA864829	Human tum
15	1000	29.4	2493	7	ADA52976	Human cod
16	882	25.9	776	5	AA92253	DNA encod
17	851	25.0	1260	5	AA87643	DNA encod
c	831	24.4	655	5	AA80955	DNA encod
18	712	20.9	437	2	AA51663	Human sec
19	689	20.2	560	5	AA91991	DNA encod
20	610	17.9	396	5	AA91986	DNA encod
21	610	17.9	396	5	AA80954	DNA encod
22	610	17.9	396	5	AA92246	DNA encod
23	610	17.9	396	5	AA87635	DNA encod
24	610	17.9	396	5	AA92247	DNA encod
25	577	17.0	498	5	AA92247	DNA encod
26	565.5	16.6	3878	7	ABT17849	Aspergill
27	565.5	16.6	4139	7	ABT19663	Aspergill
28	563	16.5	2139	7	ABT20857	Aspergill
29	563	16.5	2139	7	ABT20259	Aspergill
30	561.5	16.5	1878	7	ABT18443	Aspergill
31	556	16.3	349	3	AACT7533	Human ORF
c	550.5	16.2	585	5	AA92248	DNA encod
32	542.5	15.9	1830	7	ABT19037	Aspergill
33	505	14.8	341	2	AAV90248	EST clone
34	499.5	14.7	1671	6	ABZ32483	Candida a
35	487	14.3	2991	5	AA91992	DNA encod
36	487	14.3	2991	5	AA91992	DNA encod
37	487	14.3	2991	5	AA91992	DNA encod
38	456.5	13.4	1815	6	ABZ14399	Arabidops
39	420.5	12.4	1884	7	ACC61849	Gene sequ
40	400	11.8	1465	4	ABL03627	Drosophil
41	398	11.7	17803	4	AAK68676	Human imm
42	392	11.5	34269	4	AAK68677	Human imm
43	392	11.5	34269	4	AAK85168	Human imm
44	391.5	11.5	1293	9	ADB69802	C. neofor
45	387.5	11.4	2216	3	AAAC39853	Arabidops

ALIGNMENTS

RESULT 1  
ABV72159  
ID ABV72159 standard; DNA; 2068 BP.  
XX  
AC ABV72159;  
XX  
DT 05-DEC-2002 (first entry)  
XX  
DE Nucleotide sequence of human RRN3.  
XX  
KW Human; RNA polymerase I transcription factor;  
KW hypoproliferative disease; hyperproliferative disease; cancer;  
KW malignancy; hyperplasia; metaplasia; dysplasia; benign tumour;  
KW hyperproliferative disorder; benign dysproliferative disorder;  
KW autoimmune disease; cardiac disease; degenerative disorder;  
KW growth deficiency; hypoproliferative disorder; physical trauma; lesion;  
KW wound; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 69..2024  
FT /\*tag= a  
FT /product= "RRN3"  
XX  
XX US2002090706-A1.  
XX  
XX 11-JUL-2002.  
XX  
XX 16-AUG-2001; 2001US-00932678.  
XX  
XX 16-AUG-2000; 2000US-0225893P.  
PR

XX (REED/) REEDER R H.  
PA (MOOR/) MOOREFIELD B.  
PA (GREE/) GREENE E A.  
XX  
XX Reeder RH, Moorefield B, Greene EA;  
PI  
XX  
XX WPI; 2002-681729/73.  
DR P-PSDB; ABB78308.  
XX  
XX Novel RNN3 polypeptide, an eukaryotic RNA polymerase I transcription  
PT factor and polynucleotide encoding it useful for diagnosing, treating  
PT hyper and hypoproliferative diseases in mammals and to identify  
PT modulators.  
XX  
PS Claim 6; Page 25-26; 3lpp; English.  
XX  
XX The present sequence encodes a human RNN3 polypeptide. RNN3 is an  
CC eukaryotic RNA polymerase I transcription factor. RNN3 is useful for  
CC screening for hypo or hyperproliferative diseases, including cancer,  
CC malignancy, hyperplasia, metaplasia, dysplasia, benign tumour,  
CC hyperproliferative disorder, benign dysproliferative disease, autoimmune  
CC disease or cardiac disease. It is also used to treat diseases involving  
CC decreased cell proliferation, including degenerative disorders, growth  
CC deficiencies, hypoproliferative disorders, physical trauma, lesions and  
CC wounds. Rnn3 polypeptide, nucleic acid or antibody are also useful in  
CC diagnostics, to detect, prognose, diagnose or monitor various diseases.  
CC RNN3 nucleic acids are useful to identify other mammalian genes that  
CC encode RNN3-like molecules, to screen for mutations in a RNN3 gene that  
CC are associated with certain diseases. The polypeptide is useful as an  
CC immunogen to generate antibodies which immunospecifically bind RNN3  
CC polypeptides  
XX  
SQ Sequence 2068 BP; 573 A; 437 C; 485 G; 573 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 0 Length: 2068  
Score: 3403.00 Matches: 651  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-932-678-2 (1-651) x ABV72159 (1-2068)

QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAla 20  
DB 69 ATGGGGGCACCGTCGCTTCACACGGGTTTGGCGGGAGATGCGGCGCTTCGTCCTGCA 128  
QY 21 ValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40  
DB 129 GTTAAGAGCTGGCGCGCTCGAGGACTGGGATTTCAATATGGGTGCATTAGAGATGAC 188  
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60  
DB 189 TTTTTCAAATTCFCCCCCAAGAAAAAATGTTGCGTTTGGTGGAACTGTGACAGAGTCTTG 248  
QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
DB 249 CTGAAGTACAAAGAGGTTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAGCTGTAGAT 308  
QY 81 ProAspIleLysAspAspGlnIleLeuAsnTyrLeuLeuGluPheArgSerSerIleMet 100  
DB 309 CCAGACATAAAGAGATGACAGATCATCAACTGGCTGTAGATTCGTTCTTCATCATG 368  
QY 101 TyrLeuThrLysAspPheGluGlnIleLeuSerIleLeuLeuArgLeuProTyrLeuAsn 120  
DB 369 TACTTGACAAAAGACTTTTGAGCAACTATACGATATATATTAAGATTGCGTTGGTTGAAT 428  
QY 121 ArgSerGlnThrValValGluGluThrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
DB 429 AGAAGTCAACAGTAGTGAAGAGATATTGGCTTTTTCITGGTAATCTTGTATCAGCACAG 488

QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160  
DB 489 ACTGTTTTCCTCAGACCGGTCTCAGCATGATGCTCCATTTTGTGCTCCCGAGTG 548  
QY 161 IleIleLysGluGlyAspValAspValSerAspSerAspGluAspAspAsnLeuPro 180  
DB 549 ATCAITTAAGGAAGCGGATGATGTTTCAGATTCTGATGATGAAGATGATATCTTCCT 608  
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleAlaArgTyrValProSerThr 200  
DB 609 GCAAAATTTTGACACATGTCCAGAGCCCTGCAAAATAATAGCAAGATATCTACCATCGACA 668  
QY 201 ProTyrPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220  
DB 669 CCGTGGTTTCTCATGCCAATACTGGTGGAAAAAATTTCCATTTGTTCGAAAAATCAGAGAGA 728  
QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
DB 729 ACACCTGGAAATGTACGTTTCATAACTTACTAAGGATTAGTGATATTTTCAACCTTGAGG 788  
QY 241 HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
DB 789 CATGAATTCCTGAGCTTATTATTGAAAAAATACTACTCAAGTTGATGTGAATGCATCCCG 848  
QY 261 GlnGlyIleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
DB 849 CAGGGTATTGAAGATGCTGGAAGAAACAGCACTCAAACTTGTGGTGGGACAGATCCACG 908  
QY 281 GluGlyLeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyPro 300  
DB 909 GAAGGATTGTTTAATATGATGATGAAGTGAAGAACTGAACATGAACAAAGGCTGGTCTCT 968  
QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
DB 969 GAACGGCTCGACCATGCTGTCATCTCTGAGCGGCGCTCGACATCTCGATGCTGTTTG 1028  
QY 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340  
DB 1029 GTTTTGTCTCATATGAAGGATGTCGTATGTAGATGTTGAAGTTGATTAACGGCAAAACA 1088  
QY 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla 360  
DB 1089 AAGGATCTATATCGCGACCTGATAAACATCTTTTGACAAACTCTGTGTGCCACCCATGCC 1148  
QY 361 SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380  
DB 1149 TCCTGCCATGTACAGTTTTTCATGTTTTTACCTCTGTAGTTTCAAAATTTGGGATTCGACAG 1208  
QY 381 AlaPheLeuGluHisLeuTyrLysLysLeuGlnAspProSerAsnProAlaIleIleArg 400  
DB 1209 GCATTTTGGAAACATCTCTGGAAAAAATTCAGAGACCCAGATTAATCTGCGCATCATCAG 1268  
QY 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420  
DB 1269 CAGGCTGCTGGAAATATATATTGAAAGCTTTTGGCAAGAGCTAAATTTATCTCTTATT 1328  
QY 421 ThrValLysSerCysLeuAspLeuValAsnTyrLeuHisIleTyrLeuAsnGln 440  
DB 1329 ACTGTAAATCATGCTAGATCTTTTGGTTAACTGGCTGCATATACCTTAAATACCCAG 1388  
QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460  
DB 1389 GATTCGGGNACAAAGCATTCGCGATGTTGCTCTCCATGGACCAATTTTACTCAGCCTGC 1448  
QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
DB 1449 CAAAGCTGTGTTCTACACCTTTTGTGTTTACACAAAGCAGCTTTTGAGCGGAAACCTGAAA 1508  
QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500  
DB 1509 GAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGACGGATAGTAGATGAGCCAGCTAAAT 1568  
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr 520

```

Db 1569 CCCCTGAAGATTGCTGCCCTCAGTGGTTAACTTTTGGCTGCAATCAAAATAGTAC 1628
QY 521 GlnLeuValPheCysTyrThrIleleGluArgAsnAsnArgGlnMetLeuProValIle 540
Db 1629 CAGCTCGTCTTCTGTACACCATCATTTGAGAGGACATCGCAGATGTCGCCAGTCATT 1688
QY 541 ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560
Db 1689 AGGAGTACCGCTGGAGGAGACTCAGTCGAGATCTGCACAAACCCGCTGGACACCTTCTTC 1748
QY 561 ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580
Db 1749 CCCTTGTATCCCTGTGCTGGAAGAGTCAAGAAATTCATGTATCTTATCAGGTA 1808
QY 581 TrpGluAspMetSerAlaGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
Db 1809 TGGGAAGACATGAGTGTGAGAGCTACAGGAGTTCAGAAACCCCATGAAAGGACATA 1868
QY 601 ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620
Db 1869 GTGGAAGATGAAGATGATGATCTTCTGAAGCGAAGTGCCTCCAGAAATGATACCGTGATT 1928
QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerValGlySer 640
Db 1929 GGGATCACACCAAGCTCTTTGACAGCATTTCCGAAGTCCTTCAAGTAGTGTGGGCTCC 1988
QY 641 ProProValLeuTyrMetGlnProSerProLeu 651
Db 1989 CCACCCGTTGTATACATGCAACCCAGTCCCTC 2021
RESULT 2
AAA98384
ID AAA98384 standard; cDNA; 2040 BP.
XX
AC AAA98384;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human RNA polymerase I transcription factor TIF-1A cDNA.
XX
KW RNA polymerase I transcription factor TIF-1A; antitumor; treatment;
KW antiproliferative; cell proliferation; cancer; tissue regeneration; ss.
XX
OS Homo sapiens.
XX
PN WO200055316-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-DE000767.
XX
PR 17-MAR-1999; 99DE-01011992.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI Grummt I, Vingron M;
XX
DR WPI; 2000-587527/55.
DR P-PSDB; AAB10936.
XX
PT New DNA encoding the transcription factor TIF-1A, useful for preventing
PT or treating diseases associated with abnormal cell proliferation,
PT particularly tumors.
XX
PS Claim 1; Fig 2; 38pp; German.
XX
CC This invention describes a novel DNA sequence (I) that encodes the RNA
CC polymerase I transcription factor TIF-1A which has antitumor,
CC antiproliferative and proliferation-inducing activity. The invention also
CC describes (1) DNA (Ia) encoding a protein (II) with the biological
CC activity of TIF-1A; (2) a ribozyme (R) corresponding to (I) or (Ia) and
CC able to bind specifically to, and cleave, its transcribed RNA so as to

```

```

CC reduce or inhibit synthesis of the corresponding protein; (3) an
CC antisense RNA (AS) with binding properties similar to R; (4) an
CC expression vector that contains (I), (Ia) or sequences that encode R or
CC AS; (5) host cells containing the vectors of (4); (6) TIF-1A or (II)
CC encoded by (I) or (Ia); (7) preparation of TIF-1A or (II) by culturing
CC cells of (6); (8) ligands that bind to TIF-1A or (II); (9) antagonists
CC that weaken or block the activity of TIF-1A or (II); (10) a diagnostic
CC method for detecting abnormal TIF-1A expression; and (11) kit for
CC carrying out the method in (10). (I), and similar sequences that encode
CC proteins with equivalent activity, expression vectors containing them, or
CC their expression products are used to treat or prevent disorders
CC associated with reduced cellular proliferation, to stimulate cellular
CC proliferation, and to promote tissue regeneration, e.g. after injury or
CC radiation therapy. Ribozymes, antisense sequences directed against (I),
CC also ligands and antagonists of TIF-1A, are used to treat or prevent
CC disorders associated with excessive cellular proliferation and to inhibit
CC proliferation, especially in treatment of cancers. (I) and specific
CC ligands for TIF-1A (particularly antibodies (Ab)) are also useful for
CC diagnosis of altered TIF-1A expression by (in)direct determination of the
CC concentration, length and/or sequence of TIF-1A or its mRNA, e.g. for
CC detecting mutations. Ab can also be used for immunoprecipitation of TIF-
CC 1A and for isolation of related sequences from cDNA expression libraries.
CC (I) allows recombinant production of TIF-1A in sufficient quantities for
CC therapeutic use
XX
SQ Sequence 2040 BP; 563 A; 433 C; 474 G; 568 T; 0 U; 2 Other;

```

```

Alignment Scores:
Pred. No.: 0 Length: 2040
Score: 3391.00 Matches: 649
Percent Similarity: 99.69% Conservative: 0
Best Local Similarity: 99.69% Mismatches: 2
Query Match: 99.65% Indels: 0
DB: 3 Gaps: 0

```

US-09-932-678-2 (1-651) x AAA98384 (1-2040)

```

QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAla 20
Db 40 ATGCGCGCACCGCTGCTTCACACGCTTTGCCGGAGATGCGCGCTTCCTCTGCA 99
QY 21 VallysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40
Db 100 GTTAAGAAGCTGGCGCGCTCGAGGACTGGGATTTCAAATATCGCTGCATTAGAGAATGAC 159
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrClnValLeu 60
Db 160 TTTTTCATTTCTCCCCAAGAAAACCTGTCGGTTGGTGAACCTGTGACAGACTCTTG 219
QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80
Db 220 CTGAAGTACAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACACAGCTGTAGAT 279
QY 81 ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
Db 280 CCACACATAAAGGATGACCAGATCATCACTGGCTGTAGAAATCCCGTTCTTCTATCATG 339
QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsn 120
Db 340 TACTTGACAAAAGACTTTGACCAACTTATCAGTATTATTAAGATTGCCCTTGGTTGAAT 399
QY 121 ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140
Db 400 AGAAGTCAACACAGTAGTGAAGAGTATTGGCTTTTCTTGGTATATCTGTATCATGACACAG 459
QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160
Db 460 ACTGTTTTCTTCAGACCGTGTCTCAGCATGATTCCTCCCATTTTGTGCTCCCTCCCGAGT 519
QY 161 IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAsnLeuPro 180
Db 520 ATCATTAAAGGAAGCGATGATAGATGTTTCAGATTCTGATGATGAAGATGATTAATCTTCTCT 579

```

QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleAlaArgTyrValProSerThr 200  
DB 580 GCAAAATTTTGACACATGTACAGAGCCTTGCAAAATAATAGCAAGATATGTACCATCGACA 639  
QY 201 ProTyrPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220  
DB 640 CCGTGGTTTCTCATGCCATATCTGGTGGAAAAATTTCCATTTGTTCCGAAAAATCAGAGAGA 699  
QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
DB 700 ACACCTGGAATGTACCTTCATNACTTACTAAGGATTAGTGTATATTTCCAACTTGAGG 759  
QY 241 HisGluIleLeuGluLeuIleLeuLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
DB 760 CATGAAATTTCTGAGCTTTATTTGAAAACTACTCAAGTTGGATGTGAATGATCCCGG 819  
QY 261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
DB 820 CAGGGTATTGAGATGCTGAAGAAACAGCAACTCAAACTTGTGTGGGACAGATTCCACG 879  
QY 281 GluGlyLeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGluPro 300  
DB 880 GAAGGATTTGTTAATATGATGAAGATGAAGAACTGAACATGAAACAAAGGCTGTCTCT 939  
QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
DB 940 GAACGGCTCGACAGATGGTGCACTCTGTAGCCGAGCGCTCGACATCTCTGATGTTTGG 999  
QY 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340  
DB 1000 GTTTTCTCTACATGAAGGATGTCGTATGTAGATGGTAAAGTTGATAAGCGCAAAACA 1059  
QY 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla 360  
DB 1060 AAGGATCTATATCGCACCTGATPAAACATCTTTGACAAACTCTCTGTGCCACCCATGCC 1119  
QY 361 SerCysHisValGlnPheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380  
DB 1120 TCTTGCATGTACAGTTTTCATGTTTACCTCTGTAGTTTCAAAATGGGATTCGACAG 1179  
QY 381 AlaPheLeuGluHisLeuTyrLysLysLeuGlnAspProSerAsnProAlaIleLeuArg 400  
DB 1180 GCATTTTGGACATCTCTGAAAAAATTCGAGACCCAAAGTAACTCTGCCATCATCAGG 1239  
QY 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420  
DB 1240 CAGGCTGCTGGAATATATATGGAAGCTTTTGGCAAGAGCTAAATTTATCTCTTATT 1299  
QY 421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440  
DB 1300 ACTGTAAATCATGCCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTATATACCAG 1359  
QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460  
DB 1360 GAPTCGGGAACAAAGGCAATCTCGATGTGTCTCTCCATGGACCAATTTTACTCAGCTGC 1419  
QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
DB 1420 CAGCTGTGTCTACNCCTTTGTTTTAGACACAGCAGCTTTTGAGCGGAAACCTGAAA 1479  
QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500  
DB 1480 GAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGGCGGATAGTATGATGAGCCAGCTAAAT 1539  
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyr 520  
DB 1540 CCCTCGAAGATTTGCCCTGCCCTCAGTGGTTAACTTTTGTCTGCAATCAAAATAAGTAC 1599  
QY 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle 540  
DB 1600 CAGCTGTCTCTGTACACATCATTTGAGAGAAACAAATGCCAGATGCTGCCAGTCATT 1659  
QY 541 ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560

DB 1660 AGAGTACCGCTGGAGGAGACTCAGTCAGATCTGCACAAACCCGCTGGACACTTCTTC 1719  
QY 561 PropheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580  
DB 1720 CCGTTTGTATCCCTGTGTCTGAAGAGGTCAAAGAAATTCATTGATCTTATTCAGGTG 1779  
QY 581 TrpGluAspMetSerAlaGluLeuGlnGluPheLysLysProMetLysLysAspIle 600  
DB 1780 TGGGAAGACATCAGTGTCTGAAGAGCTACAGGAGTTCAAGAAACCCCAAGAAAGGACATA 1839  
QY 601 ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620  
DB 1840 GTGGAGATGAAGATGATGACTTCTGAAAGGCGAGTCCCCAGATGATACCGTGATT 1899  
QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerSerValGlySer 640  
DB 1900 GGGATCACCAAGCTCCTTTGACACGCATTTCCGAAAGTCCTTCAAAGTAGTGTGGGCTCC 1959  
QY 641 ProProValLeuTyrMetGlnProSerProLeu 651  
DB 1960 CCACCCGTGTGTACATGCAACCCAGTCCCTC 1992  
RESULT 3  
ABS78724  
ID ABS78724 standard; DNA; 1770 BP.  
XX  
AC ABS78724;  
XX  
DT 16-DEC-2002 (first entry)  
XX  
DE DNA encoding human NOVX3 protein.  
XX  
KW Human; NOVX; human disease; NOVX-associated disorder; cancer; addiction;  
KW Hodgkin disease; Von Hippel-Lindau syndrome; Alzheimer's disease; stroke;  
KW tubercous sclerosis; hypercalcaemia; Parkinson's disease; depression;  
KW Huntington's disease; cerebral palsy; epilepsy; Lesch-Nyhan syndrome;  
KW multiple sclerosis; ataxia-telangiectasia; leukodystrophy; anxiety; pain;  
KW obesity; Crohn's disease; osteoporosis; inflammatory bowel disease;  
KW infertility; inflammatory bowel disease; atherosclerosis; hypertension;  
KW scleroderma; haemophilia; diabetes; pancreatitis; autoimmune disease;  
KW asthma; arthritis; immunodeficiency; HIV; viral infection; neurogenesis;  
KW bacterial infection; parasitic infection; graft-versus-host disease;  
KW cell differentiation; cell proliferation; haematopoiesis; wound healing;  
KW angiogenesis; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200272770-A2.  
XX  
XX 19-SEP-2002.  
XX  
XX 08-MAR-2002; 2002WO-US007283.  
XX  
XX 08-MAR-2001; 2001US-0274281P.  
XX  
XX 09-MAR-2001; 2001US-0274849P.  
XX  
XX 12-MAR-2001; 2001US-0275235P.  
XX  
XX 13-MAR-2001; 2001US-0275579P.  
XX  
XX 13-MAR-2001; 2001US-0275601P.  
XX  
XX 14-MAR-2001; 2001US-0276000P.  
XX  
XX 20-MAR-2001; 2001US-0277239P.  
XX  
XX 20-MAR-2001; 2001US-0277327P.  
XX  
XX 20-MAR-2001; 2001US-0277338P.  
XX  
XX 21-MAR-2001; 2001US-0277791P.  
XX  
XX 22-MAR-2001; 2001US-0277833P.  
XX  
XX 23-MAR-2001; 2001US-0278152P.  
XX  
XX 26-MAR-2001; 2001US-0278894P.  
XX  
XX 27-MAR-2001; 2001US-0279036P.  
XX  
XX 28-MAR-2001; 2001US-0279344P.  
XX  
XX 30-MAR-2001; 2001US-0280233P.  
XX  
XX 02-APR-2001; 2001US-0280802P.  
XX  
XX 02-MAY-2001; 2001US-0288148P.  
XX  
XX 31-MAY-2001; 2001US-0294821P.

PR 31-OCT-2001; 2001US-0335302P.  
 PR 04-DEC-2001; 2001US-0338373P.  
 PR 07-MAR-2002; 2002US-0009466.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Spytek KA, Vernet CA, Tchernev VT, Malyankar UM, Gerlach VL;  
 PI Li L, Zerhusen BD, Patturajan M, Gusev VY, Kekuda R, Pena CEA;  
 PI Zhong M, Gangolli EA, Taupier RJ;  
 XX  
 DR WPI; 2002-713508/77.  
 DR P-PSDB; ABG97480.  
 XX  
 PT New NOVX polypeptides and polynucleotides, useful for preventing,  
 PT diagnosing or treating NOVX-associated disorders, e.g. diabetes, multiple  
 PT sclerosis, atherosclerosis, cancer, infections, osteoporosis or  
 PT Parkinson's disease.  
 XX  
 PS Claim 22; Page 98; 266pp; English.  
 XX  
 CC The present invention relates to a new polypeptide (NOVX). The NOVX  
 CC polypeptide, nucleic acid and antibody are useful in the manufacture of a  
 CC medicament for treating a syndrome associated with a human disease,  
 CC preferably a NOVX-associated disorder. The NOVX nucleic acids,  
 CC polypeptides and antibodies are useful for treating, preventing or  
 CC diagnosing diseases such as cancers, Hodgkin disease, Von Hippel-Lindau  
 CC syndrome, Alzheimer's disease, stroke, tuberculous sclerosis,  
 CC hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral  
 CC palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-  
 CC telangiectasia, leukodystrophies, addiction, anxiety, depression, pain,  
 CC obesity, Crohn's disease, osteoporosis, inflammatory bowel disease,  
 CC infertility, inflammatory bowel disease, atherosclerosis, hypertension,  
 CC scleroderma, haemophilia, diabetes, pancreatitis, autoimmune disease,  
 CC asthma, arthritis, immunodeficiencies, HIV, viral, bacterial or parasitic  
 CC infections, or graft-versus-host disease. The nucleic acids and  
 CC polypeptides may also be used as targets for the identification of small  
 CC molecules that modulate or inhibit e.g. neurogenesis, cell  
 CC differentiation, cell proliferation, haematopoiesis, wound healing and  
 CC angiogenesis, in gene therapy, in generation of antibodies that bind  
 CC immunospecifically to NOVX substances for use in therapeutic or  
 CC diagnostic methods. The nucleic acids are further used as hybridisation  
 CC probes, in chromosome mapping, tissue typing, preventive medicine, and  
 CC pharmacogenomics. The present nucleic acid sequence encodes a human NOVX  
 CC protein of the invention  
 XX  
 SQ Sequence 1770 BP; 492 A; 373 C; 413 G; 492 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,43e-296 Length: 1770  
 Score: 2839.00 Matches: 556  
 Percent Similarity: 85.41% Conservative: 0  
 Best local Similarity: 85.41% Mismatches: 5  
 Query Match: 83.43% Indels: 90  
 DB: 6 Gaps: 1  
 US-09-932-678-2 (1-651) x ABS78724 (1-1770)  
 Qy 1 MetAlaLaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAsnSerSerAla 20  
 Db 40 ATGGCGGCACCGCTGCTTCACACGGTTTGGCGGAGATGCGCGCTTCGCTCTGCA 99  
 Qy 21 VallysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40  
 Db 100 GTTAAGAGCTGGGCGGCTCGAGGACTGGGATTTCAATATGCGTGCATTAGAGATGAC 159  
 Qy 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValIleu 60  
 Db 160 TTTTTCAAATCTCCCCNAGAAAACCTGTCGGTTTGGTGGAACTGTGACAGAAGTCTTG 219  
 Qy 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
 Db 220 CTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTTGTTGAAGAACCCACTGTTAGAT 279  
 Qy 81 ProAspIleLysAspAspGlnIleAlaAsnTrpLeuLeuGluPheArgSerSerIleMet 100  
 Db 280 CCAGACATAAAGGATGACCAGATCATCACTGGCTGTAGAAATTCGGTCTTCTATCATG 339  
 Qy 101 TyrIleuThrLysAspPheGluGlnLeuIleSerIleLeuLeuArgLeuProTrpLeuAsn 120  
 Db 340 TACTTTGACAAAAGACTTTTGAGCAACTTATCATGATTATATTAAGATTGCTTGGTTGAAT 399  
 Qy 121 ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGluValSerAlaGln 140  
 Db 400 AGAAGTCAACAGTAGTGGAGAGATTTTGGCTTTCTTGGTAATCTTGTATCAGACAG 459  
 Qy 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160  
 Db 460 ACTGTTTCTCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCTCCCGAGG 519  
 Qy 161 IleIleLysGluGlyAspValAspValSerAspSerAspGluAspAsnLeuPro 180  
 Db 520 ATCATTAAGGAAGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCT 579  
 Qy 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200  
 Db 580 GCAAATTTTGACATGTCACAGAGCTTGCAAATATATAGCAAGATATGTACCATCGACA 639  
 Qy 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220  
 Db 640 CCGTGGTTCCTCATGCCAATACTCGTGGAATAATTTCCATTTGTCGAAAAATCAGAGAGA 699  
 Qy 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
 Db 700 ACCTGGAAATCTTACGTTTCACTAATCTTAAAGGATTAGTATATTTTCCAACTTGAGG 759  
 Qy 241 HisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
 Db 760 CATGAAATCTGGAGCTTATTTTGAATAACTCTCAAGCTGGAGTGAATGCATCCCGG 819  
 Qy 261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
 Db 820 CAGGGTATTGAAGATGCTGGAAGAACAGCAATCAAACTTCTGGTGGGACAGATTCACG 879  
 Qy 281 GluGlyLeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyPro 300  
 Db 880 GAAGGATGTTTAATATG----- 897  
 Qy 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
 Db 897 ----- 897  
 Qy 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340  
 Db 897 ----- 897  
 Qy 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla 360  
 Db 897 ----- 897  
 Qy 361 SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380  
 Db 898 -----GGATTCGCAGAG 909  
 Qy 381 AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg 400  
 Db 910 GCATTTTGGAAACATCTTTGGAAAAAACTTCAGGATCCAAAGTAATCTGCCATCATCAGG 969  
 Qy 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420  
 Db 970 CAGGCTCTCGAAATATATATTTGGAAGCTTTTGGCAAGAGCTAAATTTATTTCTTTATT 1029  
 Qy 421 ThrValLysSerCysLeuAspLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440  
 Db 1030 ACTGTAAACCATGCGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTTAATACAG 1089  
 Qy 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460

Db 1090 GATTCGGGAACAAAGGCATTCGCGATGTGCTCTCCATGGACCATTTTACTCAGCCTGC 1149  
QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
Db 1150 CAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAA 1209  
QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500  
Db 1210 GAAGGTTTGCAATCTCCTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAAT 1269  
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr 520  
Db 1270 CCCCTGAAGATTGCTGCGCTCAGTGGTTAACTTTTGTGCAATCACAATAAGTAC 1329  
QY 521 GlnLeuValPheCysTyrThrIleLeuGluArgAsnAsnArgGlnMetLeuProValIle 540  
Db 1330 CAGCTCGTCTTCTGTACACCATCATTTGAGAGGAAACAATCGCCAGATGCTGCCAGTCAAT 1389  
QY 541 ArgSerThrAlaGlyClyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560  
Db 1390 AGGATGACCGCTGGAGAGACACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTC 1449  
QY 561 PropheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580  
Db 1450 CCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGGTG 1509  
QY 581 TrpGluAspMetSerAlaGluLeuGlnGluPheLysLysProMetLysLysAspIle 600  
Db 1510 TGGGAAGACATCAGTGTCTGAAGAGTACAGGAGTTCAAGAAACCCATGAAAGACATA 1569  
QY 601 ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620  
Db 1570 GTGGAGATGAAGATGATGACTTTCGAAAGGGAAGTCCCCAGGAGATCCCGTGAT 1629  
QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerValGlySer 640  
Db 1630 GGGATCACACCAAGCTCCTTTGCACAGCATTTCCGAAGTCTTCAAGTAGTGTGGGCTCC 1689  
QY 641 ProProValLeuTyrMetGlnProSerProLeu 651  
Db 1690 CCACCGTGTGTGTACATGCAACCCAGTCCCTC 1722

RESULT 4

AAC76937  
ID AAC76937 standard; cDNA; 1418 BP.  
XX  
AC AAC76937;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF2492 polynucleotide sequence SEQ ID NO:4983.  
XX  
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW bone damage; cartilage damage; nocturnal haemoglobinuria; burn; wound;  
KW thrombosis; contraceptive; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200058473-A2.  
XX  
PD 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US008621.  
PF  
XX 31-MAR-1999; 99US-0127607P.  
PR 02-APR-1999; 99US-0127636P.  
PR 05-APR-1999; 99US-0127728P.  
PR 30-MAR-2000; 2000US-00540763.  
XX  
XX (CURA-) CURAGEN CORP.  
PA  
XX Shinkets RA, Leach M;  
PI  
XX WPI; 2000-602362/57.  
DR P-PSDB; AAB42728.  
DR  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease.  
XX  
PS Claim 5; Page 4157-4158; 5507pp; English.  
XX  
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;  
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
CC cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;  
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The  
CC sequences can be used for determining the presence of or predisposition  
CC to, or preventing or treating pathological conditions associated with an  
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
CC used to treat cancers, proliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive  
XX  
SQ Sequence 1418 BP; 422 A; 275 C; 318 G; 403 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 2,81e-223 Length: 1418  
Score: 2162.00 Matches: 413  
Percent Similarity: 98.34% Conservative: 2  
Best Local Similarity: 97.87% Mismatches: 6  
Query Match: 63.53% Indels: 1  
DB: 3 Gaps: 0  
  
US-09-932-678-2 (1-651) x AAC76937 (1-1418)  
QY 202 TrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArgThr 221  
Db 3 TGGTTTCTCATGCCAATCTGCTGGAAAAATTTCCATTGTTGCGAAAAATCAGAGAAACA 62  
QY 222 LeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThr-LeuArgHi 241  
Db 63 CTGGAATGTAGCTTCTATAAATCTACTAAGGATAGTGTATATATTTTCCAAACCTTGAGGCA 122  
QY 241 sGluIleLeuGluLeuIleLeuGluLysLeuLeuLysLeuAspValAsnAlaSerArgG 261  
Db 123 TGAATTTCTGGAGCTTATTATTGAAAACTACTCAAGTTGGATGTGAATGCATCCGGCA 182  
QY 261 nGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrG 281  
Db 183 GGGTATTGAAGATGCTGGAAGAAACAGCAACTCAAACTTTTGTGGGACAGATTCCACGGA 242  
QY 281 nGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProG 301  
Db 243 AGGATTGTTTAAATATGATGAAGATGAAGAAACTGAACATGAACAAAGAGCTGGTCTCTGA 302



QY 301 uArgLeuAspGlnMetValHiProValAlaGluArgLeuAspIleLeuMetSerLeuVa 321  
DB 303 ACGGCTCGACAGATGGTGATCTCTGATCCGAGCGCTGGACATCTCTGATGCTTTGGT 362  
QY 321 lleuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrly 341  
DB 363 TTTGTCCTACATGAGGATGCTCTATGTATGATGTAAGTTGATAAGCGCAACAAA 422  
QY 341 sAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSe 361  
DB 423 GGATCTATATCGGACCTGATTAACATCTTTGACAACTCTCTGTTGCCACCATGCCTC 482  
QY 361 rCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAl 381  
DB 483 CTGCCATGTACAGTCTTTTACCTCTTTTACCTCTGTAGTTTCAAATTTGGGATTCGCAGAGC 542  
QY 381 aPheLeuGluHisLeuTyrLysLeuGlnAspProSerAsnProAlaIleIleArgGl 401  
DB 543 ATTTTGGACATCTCTGGAAAAATTTGAGGACCAAGTAATCTCTGCCATCATCAGGCA 602  
QY 401 nAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleTh 421  
DB 603 GGCTGCTGAAATTAATTTGGAAGCTTTTGGCAAGAGCTAAATTTATCTCTTTATTAC 662  
QY 421 rValLysSerCysLeuAspLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAs 441  
DB 663 TGTAAATCATGCCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACAGGA 722  
QY 441 pSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGl 461  
DB 723 TTGGGAACAAGGATCTGGCATTTGCTCTCATGACCATTTTACTAGCCTGCCA 782  
QY 461 nAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGl 481  
DB 783 AGCTGTGTTTACACCTTTGTTTATAGACACAAGCAGCTTTTGGCGGAACCTCAAGA 842  
QY 481 uGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPr 501  
DB 843 AGGTTTGAGTATCTTCAGAGTCTGAATTTTGGCGGATAGTGATGAGCAGCTAAATCC 902  
QY 501 oLeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyrGl 521  
DB 903 CCTGAGATTTGCTGCCCTCAGTGGTTAATCTTTTCTGCAATCACAATAAGTACCA 962  
QY 521 nLeuValPheCysTyrThrIleGluArgAsnAsnArgGlnMetLeuProValIleAr 541  
DB 963 GCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATCGCCAGATGCTGCCAGTCATTAG 1022  
QY 541 gSerThrAlaGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhePr 561  
DB 1023 GAGTACCGCTGGAGAGACTCAGTGCAGACCTGCACAACCCACTGGACACCTTCTTCCC 1082  
QY 561 oPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnValTr 581  
DB 1083 CTTTGATCCCTGTGCTGAGAGGTCAAGAAATTCATTGATCTATTATTCAGGTGTG 1142  
QY 581 pGluAspMetSerAlaGluGluLeuGlnPheLysLysProMetLysLysAspIleVa 601  
DB 1143 GGAAGACATGAGTGTGAAGAGCTACAGAGTTCAAGAAACCCATGAAGAAAGGACATAGT 1202  
QY 601 lGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGl 621  
DB 1203 GGAAGATGAAGATGATGACTTTCTGAAAGGCGAAATTCCTCCAGAAATTAGTAGTAAGTGG 1262  
QY 621 yIle 622  
DB 1263 GGTC 1266

RESULT 5

ADD29697

ID ADD29697 standard; mRNA; 1582 BP.

XX

AC ADD29697;  
XX 15-JAN-2004 (first entry)  
DT XX Human tumour suppressor mRNA SEQ ID NO:153.  
DE ss; human; tumour suppressor; cancer; cancer; cytostatic; gene therapy.  
KW XX Homo sapiens.  
XX WO2003058201-A2.  
XX PD 17-JUL-2003.  
XX PF 31-DEC-2002; 2002WO-US041825.  
XX PR 31-DEC-2001; 2001US-0345317P.  
XX PA (QUAR-) QUARK BIOTECH INC.  
XX PA (CLEV-) CLEVELAND CLINIC FOUND.  
XX PI Feinstein E, Gudkov AV;  
XX WPI; 2003-598393/56.  
XX Diagnosing cancer comprises determining the polypeptide or polynucleotide levels e.g., hepatic lipase, in a sample from a subject, where a higher level compared to that in a subject free of cancer is indicative of cancer.  
XX Disclosure; SEQ ID NO 153; 272pp; English.  
XX The invention relates to a novel method for diagnosing a cancer in a subject. the method comprises determining, in a sample from the subject, the level of at least one polypeptide, where a higher level of the polypeptide compared to the level of the polypeptide in a subject free of cancer is indicative of cancer. The polypeptide is selected from any of the polypeptides encoded by the polynucleotides listed in the specification and polypeptides which are at least 70% homologous to the polypeptides. The method of the invention has cytostatic activity, and may have a use in gene therapy. the method is useful in identifying markers specific for one or several types of cancer, depending on the tissue origin, which may be used in numerous diagnostic and prognostic applications as well as cancer type-specific targets for therapeutic intervention. The compounds that modulate the activity of a tumour suppressor gene are useful in the treatment of cancer or as anti-cancer drugs. The present sequence represents a polynucleotide of the invention.

SQ Sequence 1582 BP; 364 A; 253 C; 296 G; 389 T; 0 U; 280 Other;

## Alignment Scores:

Pred. No.:	8.56e-206	Length:	1582
Score:	2001.00	Matches:	410
Percent Similarity:	79.34%	Conservative:	1
Best Local Similarity:	79.15%	Mismatches:	106
Query Match:	58.80%	Indels:	1
DB:	9	Gaps:	0

US-09-932-678-2 (1-651) x ADD29697 (1-1582)

QY 1 MetAlaAlaProLeuLeuHisThrArg-LeuProGlyAspAlaAlaAlaSerSerAl 20  
DB 23 ATGCGCGACCGCTGCTTCACGTTCTTGTTCGGAGATCGCGCGCTGCTCTCTGC 82  
QY 20 aValLysLysLeuGlyAlaSerArgThrGlyTleSerAsnMetArgAlaLeuGluAsnAs 40  
DB 83 AGTCAAGACCGCTGGCGCGTGGAGACTGGGATTTCAATATATGCTGATTAGAAATGA 142  
QY 40 pPheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValle 60  
DB 143 TTTTTCATATCTCCCCCAAGAAAACTGTTGCGTTTGGTGGAACTGTGACAGAAGTCTT 202  
QY 60 uLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAs 80

Db 203 GCTGAAGTACAAAAGGGTGAACAAATGACITTCAGTTGGTTGTCAGAACACGCTGTAGCA 262  
Qy 80 pProAspLeuLysAspAspGlnIleleAsnTyrLeuLeuGluPheArgSerSerIleMe 100  
Db 263 TCCAGACATAAAGGATGACAGACATCATCAACTGGCTGTAGAAATCCGTTCTCTGTGCAT 322  
Qy 100 tTyrLeuThrLysAspPheGluGlnLeuIleSerIleLeuArgLeuProTyrLeuAs 120  
Db 323 GTACTTGACAAAAGACTTTGAGCACTTATCAGTATTATATTGAGATTGCCTTGGTTGAA 382  
Qy 120 nArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGl 140  
Db 383 TAGAAGTCAACAGTAGTGAAGAGTATTGGCTTTCTTGGTAACTTTGTATCAGCATTA 442  
Qy 140 nThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProArgVa 160  
Db 443 GACTGTGTTTCTCAGACCCGTGTCTCAGCATGATTGCTTCCCAATTTTGTGCTCCCTGAGT 502  
Qy 160 lIleIleLysGluGlyAspValAspValSerAspSerAspGluAspAspAsnLeuPr 180  
Db 503 GATCATTAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCC 562  
Qy 180 oAlaAsnPheAspThrCysHisArgAlaLeuGlnIleleAlaArgTyrValProSerTh 200  
Db 563 TGCAAATTTTGCACATATCACAGAGCTTGCATAATAATAGCAAGATATGTACCATCGGC 622  
Qy 200 rProTyrPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluAr 220  
Db 623 ACCGTGTGTTTCTCATCCCAATCTCGTGGAAAAATTTCCATTTGTTGCGAAAAATCACAGAG 682  
Qy 220 gThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuAr 240  
Db 683 AACACTGGATGTGCTTCAATCAACTTACTAGGATTAGTATATATTTCCACCTTGAG 742  
Qy 240 gHisGluIleLeuGluLeuIleleGluLysLeuLeuLysLeuAspValAsnAlaSerAr 260  
Db 743 GCATGAAATCTGGAGCTTATTATTGAAAACTACTCAAGCTGATGTAATGCATCCCG 802  
Qy 260 gGlnGlyIleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerTh 280  
Db 803 GCAGGATATTGAAGATGCTGAAGAAAACAGCAAACTCAAACTTGGTGGGACAGATTCCAC 862  
Qy 280 rGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPr 300  
Db 863 CGAGAGATTCTTTAAATATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 922  
Qy 300 oGluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLe 320  
Db 923 NNN 982  
Qy 320 uValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysTh 340  
Db 983 NNN 1042  
Qy 340 rLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAl 360  
Db 1043 NNN 1102  
Qy 360 aSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGl 380  
Db 1103 NNN 1162  
Qy 380 uAlaPheLeuGluHisLeuTyrLysLysLeuGlnAspProSerAsnProAlaIleleAr 400  
Db 1163 GGCATTTTGGACATCTTTGGAAAACTTGCAGGATCCAAGTAATCCTGCCATCATCAG 1222  
Qy 400 gGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420  
Db 1223 GCAGGCTGTGGAATATATGGAACCTTTTGGCAGAGCTAAATTTATTTCTCTTAT 1282  
Qy 420 eThrValLysSerCysLeuAspLeuLeuValAsnTyrLeuHisIleTyrLeuAsnGln 440

Db 1283 TACTGTAAAAACCATGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 1342  
Qy 440 nAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCy 460  
Db 1343 GGATTGGGAAACAAAGGCATCTCGATGTTGCTCTCCATGGACCAATTTTACTCAGCCG 1402  
Qy 460 sGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLy 480  
Db 1403 CCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAGCAGCTTTTTCAGCGGAAACCTGAA 1462  
Qy 480 sGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAs 500  
Db 1463 ACAAGGTTTGCAGTATCTCTCAGAGTCTGAAATTTTTCAGCGGATAGTGATGAGCCAGCTAAA 1522  
Qy 500 nProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThr 517  
Db 1523 TCCCTTGAAGATTGCTGCCCTCAGTGGTTAACTTTTGTGCTGCAATCACA 1574  
RESULT 6  
AAS01562  
ID AAS01562 standard; cDNA; 1423 BP.  
XX AAS01562;  
XX  
XX 18-JUL-2001 (first entry)  
XX  
XX Human secretory molecule cDNA sptm #52.  
XX  
XX Human; secretory molecule; sptm; SPTM; library screening; gene therapy;  
KW cell signalling; cell proliferative disorder; atherosclerosis; cancer;  
KW immune system disorder; AIDS; neurological disorder; Alzheimer's disease;  
KW nervous system disease; mental retardation; developmental disorder;  
KW neuromuscular disorder; microarray; Incyte ID number 4287452dec; ss.  
XX  
OS Homo sapiens.  
XX  
XX W0200123558-A2.  
XX  
XX 05-APR-2001.  
XX  
XX 19-SEP-2000; 2000WO-US025610.  
XX  
XX 28-SEP-1999; 99US-0156624P.  
PR 28-SEP-1999; 99US-0156625P.  
PR 02-DEC-1999; 99US-0168611P.  
PR 02-DEC-1999; 99US-0168613P.  
PR 02-DEC-1999; 99US-0168614P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banville SC;  
PI Bratcher SR, Dufour GE, Cohen HJ, Rosen BH, Shah P, Chalup MS;  
PI Hillman JL, Jones AL, Yu JY, Greenawalt LB, Panzer SR, Roseberry AM;  
PI Wright RJ, Chen W, Liu TF, Yap PE, Stockdreher TK, Anshey S;  
PI Fong WT;  
XX  
XX WPI; 2001-258134/26.  
XX  
XX New secretory polynucleotides (SPTM) and the polypeptides they encode,  
PT for use in inducing antibodies and screening libraries of compounds.  
XX  
XX Claim 1; Page 152; 161pp; English.  
XX  
XX The present sequence for human secretory molecule cDNA sptm #52 (Incyte  
CC ID number 4287452dec) is 1 of 63 novel sptm cDNA sequences (AAS01511-  
CC AAS01573) which encode for the secretory polypeptides SPTM. The sptm  
CC polynucleotides are useful for screening a compound for effectiveness in  
CC altering expression of a target polynucleotide, where the target  
CC polynucleotide comprises sptm. Sptm is also useful in a method for  
CC assessing the toxicity of a test compound. Sptm and its fragments or  
CC complementary sequences, may be used to identify the presence of and/or  
CC determine the degree of similarity between two nucleic acid sequences.  
CC Sptm can also be used for a variety of diagnostic and therapeutic

CC purposes, e.g. diagnosing a particular condition, disease or disorder  
CC associated with cell signalling, such conditions include cell  
CC proliferative disorders such as atherosclerosis, and cancers including  
CC leukemia, an immune system disorder e.g. acquired immunodeficiency  
CC syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's  
CC disease, nutritional and metabolic disease of the nervous system, mental  
CC retardation and other developmental disorders, and muscular dystrophy and  
CC other neuromuscular disorders. Sptm can also be used to design probes  
CC useful in diagnostic assays, which may be used to monitor the progress of  
CC conditions or disorders associated with abnormal levels of expression of  
CC sptm. In addition sptm encoding SPTM may be used for somatic or germline  
CC gene therapy, for inducing antibodies, or in microarrays

XX Sequence 1423 BP; 417 A; 247 C; 320 G; 439 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 7 76e-203 Length: 1423  
Score: 1973.00 Matches: 393  
Percent Similarity: 80.53% Conservative: 0  
Best Local Similarity: 80.53% Mismatches: 4  
Query Match: 57.98% Indels: 91  
DB: 4 Gaps: 1

US-09-932-678-2 (1-651) x AAS01562 (1-1423)

Qy 31 I l e s e r A s n M e t A r g A l a L e u G l u A s n A s p P h e P h e A s n S e r P r o A r g L y s T h r V a l 50  
D b 194 A T T T C A A T A T G C G T G C A T T A G A A T G A T T T T T T C A A T T C C C C C A A G A A A A C T G T T 253  
Qy 51 A r g P h e G l y T h r V a l T h r G l u V a l L e u L e u L y s T y r L y s L y s G l y G l u T h r A s n A s p 70  
D b 254 C G G T T T C G T G A C T G T G A C A G A A G T C T T G C T G A A G T A C A A A A A G G G T G A A C A A A T G A C 313  
Qy 71 P h e G l u L e u L e u L y s A s n G l n L e u L e u A s p P r o A s p I l e L y s A s p A s p G l n I l e A s n 90  
D b 314 T T T G A G T T G T T G A A G A C C A G C T G T T A G A T C C A G A C A T A A A G G A T G A C C A G A T C A T C A A C 373  
Qy 91 T r p L e u L e u G l u P h e A r g S e r S e r I l e M e t T y r L e u T h r L y s A s p P h e G l u G l n L e u I l e 110  
D b 374 T G G C T G C T A G A A T T C C G T T C T T C A T C A T G T A C T T G A C A A A A G A G C T T T G A G C A A C T T A T C 433  
Qy 111 S e r I l e I l e L e u A r g L e u P r o T r p L e u A s n A r g S e r G l n T h r V a l V a l G l u G l u T y r L e u 130  
D b 434 A G T A T T A T A T T A G A T T G C C T T G T T G A A T A G A A G T C A A A C A G A G T A G T G G A A G A G A T A T T T G 493  
Qy 131 A l a P h e L e u G l y A s n L e u V a l S e r A l a G l n T h r V a l P h e L e u A r g P r o C y s L e u S e r M e t 150  
D b 494 G C T T T C T T G G T A T C T T G T A T C A G C A G A C A G A C T G T T T C C T C A G A C C G T G T C T C A G A T G 553  
Qy 151 I l e A l a S e r H i s P h e V a l P r o A r g V a l I l e L y s L y s G l u G l y A s p V a l A s p V a l S e r 170  
D b 554 A T T G C T T C C C A T T T T G T G C C T C C C G A G T G A T C A T T A A G A A G C G A T G T A G A T G T T T C A 613  
Qy 171 A s p S e r A s p G l u A s p A s n L e u P r o A l a S n P h e A s p T h r C y s H i s A r g A l a L e u 190  
D b 614 G A T T C T G A T G A T G A A G A T G A T A A T C T T C C T G C A A A T T T G A C A C A T G T C A C A G A C C C T T G 673  
Qy 191 G l n I l e L e u A l a A r g T y r V a l P r o S e r T h r P r o T r p P h e L e u M e t P r o I l e L e u V a l G l u 210  
D b 674 C A A T A T A G C A A G A T A T G T A C C A T C G A C A C C G T G T T C T A T C C C A A T A C T A C T G T G T G A A 733  
Qy 211 L y s P h e P r o P h e V a l A r g L y s S e r G l u A r g T h r L e u G l u C y s T y r V a l H i s A s n L e u L e u 230  
D b 734 A A A T T T C C A T T T G T T C G A A A A T C A G A G A A C A C T G G A A T G T T A C G T T C A T A A C T T A C T A 793  
Qy 231 A r g L e u S e r V a l T y r P h e P r o T h r L e u A r g H i s G l u I l e L e u L e u L e u I l e G l u L y s 250  
D b 794 A G G A T T A G T G T A T A T T T C C A A C C T T G A G G A T G A A A T T C T G G A G C T A T T A T T T G A A A A A 853  
Qy 251 L e u L e u L y s L e u A s p V a l A s n A l a S e r A r g G l n G l y I l e L e u A s p A l a G l u G l u T h r A l a 270  
D b 854 C T A C T C A A G T T G G A T G A T G C A T C C C G C A G G G T A T T G A A G A T G C T G A A A A A C A G C A 913

Qy 271 T h r G l n T h r C y s G l y T h r A s p S e r T h r G l u G l y L e u P h e A s n M e t A s p G l u A s p G l u 290  
D b 914 A A T C A A A C T T G T G T G G A C A G A T T C C A C G A A G G A T T G T T A A T A T G ----- 961  
Qy 291 G l u T h r G l u H i s G l u T h r L y s A l a G l y P r o G l u A r g L e u A s p G l n M e t V a l H i s P r o V a l 310  
D b 961 ----- 961  
Qy 311 A l a G l u A r g L e u A s p I l e L e u M e t S e r L e u V a l L e u S e r T y r M e t L y s A s p V a l C y s T y r 330  
D b 961 ----- 961  
Qy 331 V a l A s p G l y L y s V a l A s p A s n G l y L y s T h r L y s A s p L e u T y r A r g A s p L e u I l e A s n I l e 350  
D b 961 ----- 961  
Qy 351 P h e A s p L y s L e u L e u L e u P r o T h r H i s A l a S e r C y s H i s V a l G l n P h e M e t P h e T y r 370  
D b 961 ----- 961  
Qy 371 L e u C y s S e r P h e L y s L e u G l y P h e A l a G l u A l a - P h e L e u G l u H i s L e u T r p L y s L y s L e 390  
D b 962 ----- G G A T T C G C A G A G C C C A T T T T T G G A A C A T C T T T G G A A A A A C T T 1003  
Qy 390 u G l n A s p P r o S e r A s n P r o A l a I l e L e u A r g G l n A l a A l a G l y A s n T y r I l e G l y S e r P h 410  
D b 1004 G C A G G A T C C A A G T A A T C C T G C C A T C A T C A G C A G G C T G C T G G A A A T T A T A T T G G A A G C T T 1063  
Qy 410 e L e u A l a A r g A l a S p h e I l e P r o L e u I l e T h r V a l L y s S e r C y s L e u A s p L e u L e u V a 430  
D b 1064 T T T G G C A A G A G C T A A A T T T A T T T C T T A T T A C T G T A A A A C C A T G C C T A G A T C T T T T G G T 1123  
Qy 430 l a s n T r p L e u H i s I l e T y r L e u A s n A s n G l n A s p S e r G l y T h r L y s A l a P h e C y s A s p V a 450  
D b 1124 T A A C T G C T C C A C A T A T A C C T T A A T A C C A G A A T C G G A A C A A G C A T T C T C G A T G T 1183  
Qy 450 l a l a L e u H i s G l y P r o P h e T y r S e r A l a C y s G l n A l a V a l P h e T y r T h r P h e V a l P h e A r 470  
D b 1184 T G C T C T C A T G G A C C A T T T T A C T C A G C C T G C C A A G C T G T G T T C A C A C C T T T G T T T T A G 1243  
Qy 470 g H i s L y s G l n L e u L e u S e r G l y A s n L e u L y s G l u G l y L e u G l n T y r L e u G l n S e r L e u A s 490  
D b 1244 A C A A C A C A C C T T T T G A G C G A A A C C T G A A A G A A G G T T C A G A T A T C T T C A G A G C T G A A 1303  
Qy 490 n P h e G l u A r g I l e V a l M e t S e r G l n L e u A s n P r o L e u L y s I l e C y s L e u P r o S e r V a l V a 510  
D b 1304 T T T T G A C G G A T A G T A G A G C C A G C T A A T C C C T G A A G A T T T G C C T C A G C C T C A G T G G T 1363  
Qy 510 l a s n P h e P h e A l a A l a I l e T h r 517  
D b 1364 T A A C T T T T T T G T G C A A T C A C A 1385  
RESULT 7  
AAS44584  
ID AAS44584 standard; DNA; 1498 BP.  
XX AC AAS44584;  
XX XX  
XX 18-DEC-2001 (first entry)  
XX XX  
XX Human full-length polynucleotide sequence #9.  
XX Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;  
XX mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;  
XX cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;  
XX nervous system disorder; inflammatory disorder; cell differentiation; ds;  
XX angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;  
XX genetic disorder; bone regeneration; tendon; ligament; tissue repair;  
XX cytostatic; antirheumatic; antiarthritic; vulnery; antiparkinsonian;  
XX antibacterial; immunosuppressive; vasotropic; antiparkinsonian;  
XX neuroprotective; osteopathic; antidiabetic; antiallergic;  
XX immunostimulant; analgesic; gene therapy.

```
OS Homo sapiens.
XX WO200164834-A2.
XX 07-SEP-2001.
XX 26-FEB-2001; 2001WO-US004926.
XX 28-FEB-2000; 2000US-00515126.
XX 18-MAY-2000; 2000US-00577409.
XX 17-JUN-2000; 2000US-00597707.
XX 14-JUL-2000; 2000US-00616807.
XX 19-SEP-2000; 2000US-00664641.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Yue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
XX Drmanac R;
XX WPI; 2001-589862/66.
XX P-PSDB; AAU27684.
XX Novel polypeptides and nucleic acids obtained from cDNA libraries
XX prepared from various human tissues, for diagnosis, treatment of cancer,
XX neurological, inflammatory disorders and for use in arrays for detection.
XX Claim 1; SEQ ID NO 9; 153pp; English.
XX Sequences AAS44576-AAS44919 represent full-length polynucleotides and
XX contig polynucleotides encoding polypeptides of the invention. The DNA
XX and protein sequences are useful for the treatment, diagnosis and
XX prevention of various types of disorder in a mammalian subject such as a
XX human, dog, monkey, mouse, hamster or rat. The disorders include cancers
XX such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
XX as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
XX diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
XX disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
XX chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
XX Wernicke disease, inflammatory disorders such as nephritis. Crohn's
XX disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
XX bowel disease. The sequences exhibit activity relating to angiogenesis,
XX cell proliferation, cell differentiation, stem cell growth factor,
XX activin or inhibin. Therefore, they can be used to manipulate stem cells
XX in culture to give rise to neuroepithelial cells that can be used to
XX augment or replace cells damaged by illness, accidental damage or genetic
XX disorders. The sequences may also be used for regeneration of bone,
XX cartilage, tendons and ligaments and in tissue repair and burn healing.
XX Note: Some sequences for this patent did not form part of the printed
XX specification, but were obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1498 BP; 417 A; 332 C; 325 G; 424 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Fred. No.: 4.3e-180 Length: 1498
XX Score: 1763.00 Matches: 339
XX Percent Similarity: 97.98% Conservative: 1
XX Best Local Similarity: 97.69% Mismatches: 7
XX Query Match: 51.81% Indels: 1
XX DB: 4 Gaps: 0
XX
XX US-09-932-678-2 (1-651) x AAS44584 (1-1498)
XX
XX 223 GluCysTyrValHisAsnLeuArgIleSerValTyrPheProThrLeuArgHisGlu 242
XX 4 GAAAGTACGTTCAATCTTACAGAGGATTAGTGTATATTTCCAACTG-AGGCATGAA 62
XX
XX 243 IleLeuGluLeuIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGly 262
XX 63 ATTCGTAGGCTTATTATGAAATACTACTCAAGTTGGATGTGAATGATCCCGCAGGCT 122
XX
XX 263 IleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGly 282
XX
123 ATTGAAGATGCTGAAGAAACAGCAACTCAAACCTTGTTGGTGACAGATTCACGGAAGA 182
283 LeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyProGluArg 302
183 TTGTTTAATATGATGAAGATGAAGAAACTGAACATGAACAAAGAGCTGCTCCTGAACG 242
303 LeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu 322
243 CTCGACCAAGATGTCATCTCTAGCGAGCCCTGGACATCCTGATGCTTTGGTTTG 302
323 SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAsp 342
303 TCCTACATGAAGGATGTCGTATGTAGATGCTAAGTTGATTAACGCAAAACAAAGAT 362
343 LeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCys 362
363 CTATATCGGACCTGATTAACATCTTTGACAAACTCCTGTTGCCCAACCATCCTCTGC 422
363 HisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPhe 382
423 CATGTACAGTTTTTTCATGTTTACCTCTGTAGTTTCAAAATGGGATTCGACAGGCA 482
383 LeuGluHisLeuTyrLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAla 402
483 TTGGAACATCTCTGGAATAAATTGACGACCAAGTAATCCTGCCATCATCAGCAGGCT 542
403 AlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrVal 422
543 GCTGAAATATATTTGGAAGCTTTTGGCAAGAGCTAAATTTATTCCTCTTATTACTGA 602
423 LysSerCysLeuAspLeuValAsnTyrPheLeuHisIleTyrLeuAsnAsnGlnAspSer 442
603 AAATCATGCTTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACAGGATTG 662
443 GlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla 462
663 GGAACAAAGGCATCTCGCATGTTGCTCTCCATGACCATTTTACTAGCCTGCCAAGCT 722
463 ValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482
723 GTGTTCTACACCTTTGTTTATAGACAAAGCAGCTTTTGGAGCGGAAACCTGAAAGAAG 782
483 LeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeu 502
783 TTGAGTATCTTCAGAGCTCTGAATTTGAGCGGATAGTGTAGTACGACCACTAAATCCC 842
503 LysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeu 522
843 AAGATTTGCCCTGCCCTCAGTGGTTAACTTTTGTGCAATCACAAATAAGTACCAGCTC 902
523 ValPheCysTyrThrIleLeuArgAsnArgGlnMetLeuProValIleArgSer 542
903 GTCCTTCTACACCATCATCGAGAGGAACAAATCGCCAGATGCTGCCAGTCATTAGGAGT 962
543 ThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPhe 562
963 ACCACTGGAGGAGACTCAGTGCAGACCTGCACAAACCCACTGGACACCTTGTCCCTTT 1022
563 AspProCysValLeuLysArg 569
1023 GATCCCTGTGTGCTGAAGAGA 1043
XX
XX RESULT 8
XX AAS87641
XX ID AAS87641 standard; cDNA; 3169 BP.
XX AC AAS87641;
XX XX
XX DT 13-FEB-2002 (first entry)
XX XX
XX DE DNA encoding novel human diagnostic protein #23445.
XX XX
```

KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US008631.  
 XX PR 31-MAR-2000; 2000US-00540217.  
 XX PR 23-AUG-2000; 2000US-00649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX DR WPI, 2001-639362/73.  
 XX DR P-PSDB; ABG23454.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.  
 PT  
 PT  
 XX Claim 1; SEQ ID NO 23445; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 3169 BP; 812 A; 841 C; 779 G; 737 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 6.27e-131 Length: 3169  
 Score: 1313.00 Matches: 293  
 Percent Similarity: 63.45% Conservative: 23  
 Best Local Similarity: 58.84% Mismatches: 57  
 Query Match: 38.58% Indels: 125  
 DB: 5 Gaps: 9

US-09-932-678-2 (1-651) x AAS87641 (1-3169)

QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAla 20  
 Db 1 ATGGCGGACCGCTGCTTCACAGCGTTTGGCGGAGATGCGCGCTTCGCCCTCTGCA 60  
 QY 21 ValLysLeuLeuGlyAlaSerArgThrGlyLeSerAsnMetArgAlaLeuGluAsnAsp 40  
 Db 61 CTCAGATGCTGGGCGCGCTCGAGGACTCGGATTTCAATATCGCGTCAATAGAAATGAT 120  
 QY 41 PhePheAsnSerProArgLysThrValArgPheGlyThrValThrGluValLeu 60  
 Db 121 TTTTCAATTCTCCCCCAAGAAACTGTTGAGTTGGTGAACACTGTGACAGAAGTCCTG 180

QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
 Db 181 CTGAAGTACAAACGGGTGAACAACAAAGACCTTGAAGTTGTGAAGAACACAGCTGTATAT 240  
 QY 81 ProAspLeuLys-----AspAspGlnLeuLeuAsnTrpLeuLeuGluPheArgSerSer 98  
 Db 241 CCAGACATAAAGGTGTGTGTCGCCCGCATGCGCATGATCTCTGAGGAGGAGGAGGCCCAT 300  
 QY 99 Ile-----MetTyrLeuThrLysAspPheGluGln 108  
 Db 301 GTGGCGAGTGTGCTTTCATCCCTTCGTGGCTTCAGTTGTGCTCAACACGAGAGTGGC 360  
 QY 109 LeuIleSerIleLeuLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGlu 128  
 Db 361 CTCAGCCCGAAGTTGCTAAGCTGCTCGGTGGGTGG----- 399  
 QY 129 TyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeu 148  
 Db 400 -----CCCGTGGCTGCGCGGCCACATAC 423  
 QY 149 SerMetIleAlaSerHisPheValProProArgValIleLeuLysGluGlyAspValAsp 168  
 Db 424 AGCCCCAGCCTCTGCCAC----- 441  
 QY 169 ValSerAspSerAspGluAspAsnLeuProAla-AsnPheAspThrCysHisAr 188  
 Db 442 -----AACCCCCAACCCGAGCGAGCTTCGGCTCCTCAT-- 478  
 QY 188 gAlaLeuGlnIleLeuAlaArgTyrValProSerThrTrpPheLeuMetProIleLe 208  
 Db 479 -----TGCTTTGCGAG-----ACCGTGGTTTCTCATGCCAATACT 516  
 QY 208 uValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAs 228  
 Db 517 GGTGGAATAATTCATTTGTTGGAATAATCAGAGAAACCTGGAATTTACGTTCAATA 576  
 QY 228 nLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIle 248  
 Db 577 CTTACTAAGATTAGTGTATATTTTCCACCTTGAGGCATGAAATCTGGAGCTATTAT 636  
 QY 248 eGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyLeuGluAspAlaGluG 268  
 Db 637 TGAATAAATCTACTCAAGTTGGATGTAATGCATCCCGGAGGATTTGAAGATGCTGAAGA 696  
 QY 268 uThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspG 288  
 Db 697 AACAGCAATCAAACTTGTGTGGGACAGATTCACCGAAGGATTTGTGTAATATGTTAG 756  
 QY 288 uAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHi 308  
 Db 757 C----- 757  
 QY 308 sProValAlaGluArgLeuAspIleLeuMetSerSerLeuValLeuSerTyrMetLysAspVa 328  
 Db 758 -----AGTTTATTAATGAAAGTGGAGATGAAGTTTATC----- 790  
 QY 328 lCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIl 348  
 Db 791 -----ATAAT 795  
 QY 348 eAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMe 368  
 Db 796 CAAGGGGTGGAACACAGCTAGTGTG----- 820  
 QY 368 tPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLy 388  
 Db 821 -----CTCATCTTTGGATTGCGAGAGCATTTTGGAAACCTCTTTGGA 864  
 QY 388 sLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGl 408  
 Db 865 AAAATTGAGGATCCAAGTAACTCTGCCATCATCAGGAGGCTGCTGGAATTATATTTGG 924

QY 408 ySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLe 428  
Db 925 AAGCTTTTGGCAGAGCTAAATTTATCTCTTATTACTGTAAATCAATGCTAGACT 984  
QY 428 uLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCy 448  
Db 985 TTTGGTTAACTGGCTGCACATATACCTTAATAACCAAGGATTCGGGAACAAGGCATTCTG 1044  
QY 448 sAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheVa 468  
Db 1045 CGATGTTGCTCTCCATGGACCAATTTACTAGCCTGCCAAGCTGTGTTCTACACCTTGT 1104  
QY 468 lPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyr 485  
Db 1105 TTTTAGACACAGCAGCTTTTGAGCGGAAACCTGAAAGAGGGCTCTGGTTC 1156  
RESULT 9  
ABK47950  
ID ABK47950 standard; cDNA; 1989 BP.  
XX  
AC ABK47950;  
XX  
XX  
DT 02-JUL-2002 (first entry)  
XX  
DE Human transcriptional control factor ZFM1 isomer 25.63 cDNA.  
XX  
KW Human; transcriptional control factor ZFM1 isomer 25.63; gene; ss; HIV;  
KW malignant tumour; haemopathy; human immunodeficiency virus; cancer;  
KW immunological disease; inflammation; cytostatic; haemostatic; virucide;  
KW immunomodulatory; antiinflammatory; gene therapy.  
XX  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 1178..1879  
FT /\*tag= a  
FT /product= "Human transcriptional control factor ZFM1  
FT isomer 25.63"  
XX  
XX  
PN WO20020588-A1.  
XX  
XX  
PD 14-MAR-2002.  
XX  
XX  
PF 02-JUL-2001; 2001WO-CN001127.  
XX  
XX  
PR 07-JUL-2000; 2000CN-00117050.  
XX  
XX (BIOW-) BLOWDOWN GENE DEV INC SHANGHAI.  
XX  
XX  
PI Mao Y, Xie Y;  
XX  
XX  
DR WPI: 2002-339796/37.  
XX  
XX P-PSDB; AAU77171.  
XX  
XX  
PT Human transcriptional control factor ZFM1 isomer 25.63 and encoding  
PT polynucleotide, used in diagnosis and treatment of malignant tumors,  
PT hemopathy, human immunodeficiency virus infection, immunological diseases  
PT and inflammation.  
XX  
XX  
PS Claim 6; Page 30-31; 38pp; Chinese.  
XX  
XX  
CC The invention relates to the human transcriptional control factor ZFM1  
CC isomer 25.63 and the polynucleotide encoding it. The sequences of the  
CC invention are used in diagnosis and treatment of malignant tumours,  
CC haemopathy, human immunodeficiency virus (HIV) infection, immunological  
CC diseases and various inflammations. This sequence represents cDNA  
CC encoding the human transcriptional control factor ZFM1 isomer 25.63  
XX  
SQ Sequence 1989 BP; 559 A; 356 C; 430 G; 644 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1e-127 Length: 1989  
Score: 1280.50 Matches: 305

Percent Similarity: 40.90% Conservative: 5  
Best Local Similarity: 40.24% Mismatches: 21  
Query Match: 37.63% Indels: 428  
DB: 6 Gaps: 5  
US-09-932-678-2 (1-651) x ABK47950 (1-1989)  
QY 126 ValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArg 145  
Db 1 GTGGAGAGATATTGGCTTTTCTTGGTAATCTTGATCAGCATAGACTGTTTTCCTCAGA 60  
QY 146 ProCysLeuSerMetIleAlaSerHisPheValProProArgValIleIleLysGluGly 165  
Db 61 CGGTGTCTCAGCATGATTGCTTCCCATTTTGTGCTCCCGAGTGATCATTAAGGAAGGC 120  
QY 166 AspValAspValSerAspSerAspGluAspAspAsnLeuProAlaAsnPheAspThr 185  
Db 121 GATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCTCGCAAAATTTGACACA 180  
QY 186 CysHisArgAlaLeuGlnIleIleAlaArgTyrValProSer- 199  
Db 181 TATCACAGAGCCTTGCAATAATATAGCAAGGTATGACCATCATGAGTATACATTTTCTTCA 240  
QY 199 ----- 199  
Db 241 TTTTGAATGTTTAATTTCTCAGAGAAAATTGTAATCAATTAGTAAAAAATTATAAAATCTTAA 300  
QY 199 ----- 199  
Db 301 TAGTATTAAAGCTTGAGTCTTACATTTGCATGTTTTTTTTTTTGTATPCCACTTGAGGAAAC 360  
QY 199 ----- 199  
Db 361 ATTACATCTCACAAAAGTGGCAATTTCCATTTTCTATTATTCTCTTTAACTGTTTTTCA 420  
QY 199 ----- 199  
Db 421 AAGTTCTGATGATGAGATTCTCCCCCAATTTTGTATGGTGGTGGAAATTTGTTTTTATCTT 480  
QY 199 ----- 199  
Db 481 CAACAGATATGCTATCCAAAATTTTTCAGTGAAGAAACCCCTGGGTGTGTGTGTGTCATGC 540  
QY 199 ----- 199  
Db 541 CATATGAATAAAAATTGCACITCTAAGAAAAGCTTTTCAGGTTTGTTGGGTTTCTTTTGA 600  
QY 199 ----- 199  
Db 601 GGGGTGGATTCTAGTTCCCTCTGCTGTGTGATTATTGTTAACTTAAAAAATCCAACCT 660  
QY 199 ----- 199  
Db 661 TGAATTATTTTTTCTTCTTTTAAAAATAATATACATGTGTAGTGGGAAATGTGACGAAAAG 720  
QY 199 ----- 199  
Db 721 TGCCTGTATGTTCTCTGGGAGAGAGCTCCCTCTTGTGATTGCTGTTGATATCAGAGTT 780  
QY 199 ----- 199  
Db 781 AACAGAAGCTTATTTTCTCTAAGTTGTTTATAGACTTTTCTCAGAAGCTATACATTTGTAAGT 840  
QY 199 ----- 199  
Db 841 TCCAGTTCTGGCGGGGGGGTGGCTCACACCTGTAACTCCAGCACTTTGGGAGGCTGAG 900  
QY 199 ----- 199  
Db 901 GCGGGGGGATCACCTGAGGTGCGGAGTTCGAGACTAGCCTGACCAACATCGAGAAACCCC 960  
QY 199 ----- 199

Db 961 GTCTCTACTAAAAATACAAATTAGTGGCGGTGGTGGCGCATGCTGTAATCCAGCTG 1020  
QY 199 ----- 199  
Db 1021 TTTGGGGGCTGAGGCAGGAGAATCGTTGAACCCGGGAGCAGAGGTTACAGTGAGCGG 1080  
QY 199 ----- 199  
Db 1081 AGATTGTGCACCTGCACCTCAGCTGGGCAACAAGAGTGAATCCGTCCTCAAAAACAA 1140  
QY 200 -----ThrProTyrPheLeuMetProIleLeuValGluLysPh 212  
Db 1141 AAAAAAGTTCAGTTCCTTTGAGGGCACCGTGGTTTCTCATGCCAATACTGGTGGAAAAAT 1200  
QY 212 eProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgI 232  
Db 1201 TCCATTTGTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTTCATACTTAAGGAT 1260  
QY 232 eSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleGluLysLeuLe 252  
Db 1261 TAGTGTATATTTTCCAACTTGAGGCATGAATTCGGAGCTTATTATTGAAAACTACT 1320  
QY 252 ubLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluThrAlaThrGl 272  
Db 1321 CAAGCTGGATGTGAATGCATCCGGCAGCGGTATTGAAGATGCTGAAGAAACAGCAATCA 1380  
QY 272 nThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluTh 292  
Db 1381 AACTTTGTGTGGGACAGATTCCACGGAAGGATGTTTAAATAG----- 1423  
QY 292 rGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGl 312  
Db 1423 ----- 1423  
QY 312 uArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAs 332  
Db 1423 ----- 1423  
QY 332 pGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAs 352  
Db 1423 ----- 1423  
QY 352 pLysLeuLeuProThrHisAlaSerCysHisValGlnPheMetPheTyrLeuCy 372  
Db 1423 ----- 1423  
QY 372 sSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAs 392  
Db 1424 -----GGATTTCGACAGAGGCATTTTGGAAACATCTTTGGAAAAACTTCGAGGA 1470  
QY 392 pProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAl 412  
Db 1471 TCCAAAGTAATCTCGCATCATCAGCAGCGCTGCTGGAATAATATATATGGAAGCTTTTGGC 1530  
QY 412 aArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTr 432  
Db 1531 AAGAGCTAAATTTATTTCTTTATCTATCTAGCAAAACCATGCTAGATCTTTGGTTAACTG 1590  
QY 432 pLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLe 452  
Db 1591 GCTGCACATATACCTTAATAACAGGATTCGGGAACAAAGCAATTCGCGATGTGTCTCT 1650  
QY 452 uHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLy 472  
Db 1651 CCATGGACCATTTTACTCAGCGCTGCAAGCTGTGTTCTACACCTTTGTTTTAGACACAA 1710  
QY 472 sGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGl 492  
Db 1711 GCAGCTTTTGAGCGGAACCTTGAAGAAGAGGTTTCAGGTATCCTCAGAGTCGGAATTTTGA 1770  
QY 492 uArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPh 512  
Db 1771 GCGGATAGTAGACCCAGCTAAATCCCTCAGAGATTTCCTGCCCTCAGTGGTTAACTT 1830

QY 512 ePheAlaAlaIleThrAsnLysTyr-GlnLeuValPheCysTyrThrIleIleGluArgA 532  
Db 1831 TTTTGTGCAATCACA--AAGATGAAGACTTGTG----- 1862  
QY 532 snAsnArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleC 552  
Db 1863 -----GATATGGATGGTGGTGGTGT-----T 1884  
QY 552 yThrAsnProLeuAspThrPhePheProPheAspProCysValleuLys 568  
Db 1885 GCACAAACAAT-ATCAATTTATTTTATACCACTGAACCGGCACTTCAAAA 1933  
RESULT 10  
AAS92254  
ID AAS92254 standard; cDNA; 3175 BP.  
XX AAS92254;  
AC AAS92254;  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #28058.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG28067.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 28058; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 3175 BP; 811 A; 849 C; 779 G; 736 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 3,01e-119 Length: 3175  
Score: 1205.00 Matches: 288  
Percent Similarity: 61.90% Conservative: 24  
Best Local Similarity: 57.14% Mismatches: 61  
Query Match: 35.41% Indels: 131  
DB: 5 Gaps: 9

US-09-932-678-2 (1-651) x AAS92254 (1-3175)

```
Qy 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaSerSerAla 20
Db 1 ATGGCGGCGCGCTGCTTACACGCGTTTGCAGGAGATCGCGCCCTTCGCGCTTGCA 60
Qy 21 ValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40
Db 61 GTCAAGATGCTGGCGCGTGCAGGACTGGGATTTCAAAATATGCGTCATTAGAGAATGAT 120
Qy 41 PhePheAsnSerProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60
Db 121 TTTTTCAAATCTCCCGCAAGAAACTGTTCAGTTTGGTGGAACTGTGACAGAAAGTCTTG 180
Qy 61 LeuLysTyrlLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80
Db 181 CTGAGGTACMAACGGGTGAACAAATGACTTTGAGTTGTGAAGAACCCAGCTGTAGAT 240
Qy 81 ProAspIleLys-----AspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSer 98
Db 241 CCAGACATAAAGTGTGTGTCGCCCGCATGCGCATGATCTCTGAGGGAGGGGGCCATT 300
Qy 99 Ile-----MetTyrlLeuThrLysAspPheGluGln 108
Db 301 GTGGGCGAGTGTGCTTTCATCCCTTCGTGGCTTTCAGTTGCTCAACACGAGTGGC 360
Qy 109 LeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGlu 128
Db 361 CTCAGCCGAGGTTGCTAAGCTGCTGCGTGGTGGTGG----- 399
Qy 129 TyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeu 148
Db 400 -----CCCGTGGCGCTCGCGGCCACATAC 423
Qy 149 SerMetIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspValAsp 168
Db 424 AGCCCCAGCGCTCTGCCAC----- 441
Qy 169 ValSerAspSerAspGluAspGluAspAsnLeuProAla-AsnPheAspThrCysHisAr 188
Db 442 -----AACCCCAACGCCCGCAGCTTCTGGCTCACTCAT-- 478
Qy 188 gAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProIleLe 208
Db 479 -----CTGCTTTCAGG-----ACACCGCTGGTTTCTCATGCCAATACT 516
Qy 208 uValGluLysPheProPheValArgLysSerGluArgqThrLeuGluCysTyrValHisAs 228
Db 517 GGTGGAAAAATTTCCATTTGTCGAAAATCAGAGAGAACACTGGAATGTTACGTTTCATAA 576
Qy 228 nLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuLeuIleI 248
Db 577 CTTACTAAGGATTAGTGTATATTTCACACTTGAGGCATGAAATTCGGAGCTTATAT 636
Qy 248 eGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGl 268
Db 637 TGAATAAACTACTCAAGTTGGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGA 696
Qy 268 uThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGl 288
Db 697 AACAGCAATCAAACTTGTGGTGGACAGATTCCACGGAGGAGTTGTTAATATGTTAG 756
Qy 288 uAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHi 308
```

```
Db 757 C----- 757
Qy 308 sProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspVa 328
Db 758 -----AGTTTATTATGAAGTGGAGATCAAGTTTATC----- 790
Qy 328 lCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIl 348
Db 791 -----ATAAT 795
Qy 348 eAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMe 368
Db 796 CAAGGGTGGAAACAGCTAGTGTG----- 820
Qy 368 tPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLy 388
Db 821 -----CTCATCTTTGGATTTCGAGAGGCATTTTGGAAACCTCTTTGGAA 864
Qy 388 sLysLeuGln-AspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleG 408
Db 865 AAAACTGCAGGGATCCAGTAATCTGCCATCATCAGGCAGGCTGCTGGACATTATATTG 924
Qy 408 lYserPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspL 428
Db 925 CRAAGCTTTTGGCAAGAGCTACATTTATTCTTATTACTGTAAAAACCATGCTAGATC 984
Qy 428 euleuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSer-GlyThrLysAlaPhe 447
Db 985 TTTTGGTTAACTGGCTGCATATATACCTTTAATAACAGGATTCGGGAAACAAAGGCATTC 1044
Qy 448 CysAspValAla-LeuHisGly-ProPheTyrSerAlaCys-GlnAlaValPheTyrThr 466
Db 1045 TGGCATGTTGCCCTCTCCATGGACCCATTTTACTCAGCTGCCCAAGCTGTGTTCTACAC 1104
Qy 467 PheValPheArg-HisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyr 485
Db 1105 TTTGTTTTTAGAACACACAGCAGCTTTTGACGGGAAACCTGAAAGAGCGCTCTGTTTC 1162
```

## RESULT 11

AAS87636  
ID AAS87636 standard; cDNA; 1017 BP.

XX AC AAS87636;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #23440.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX XX Drmanac RT, Liu C, Tang YT;

XX XX WPI; 2001-639362/73.

XX XX P-PSDB; ABG23449.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess



biodiversity.

Claim 1; SEQ ID NO 23440; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal genes. (I) is useful in gene disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AA564197-AA594564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct/sequences](http://www.int/pub/published/pct/sequences)

4121  
SQ Sequence 1017 BP; 301 A; 214 C; 230 G; 272 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.06e-117	Length:	1017
Score:	1181.00	Matches:	249
Percent Similarity:	55.31%	Conservative:	1
Best Local Similarity:	55.03%	Mismatches:	4
Query Match:	34.70%	Indels:	198
DB:	5	Gaps:	2

US-09-932-678-2 (1-651) x AAS87636 (1-1017)

QY	30	GlyIleSerAsnMetArgAlaLeuGluAsnAspPheAsnSerProProArgLysThr	49
Dd	7	GGGAATTCAAATATCGTGCATTAGAGAATGATTTTCAATTCTCCCCAAGAAAAC	66
QY	50	ValArgPheGlyGlyThrValThrGluValLeuLeuLysTyrlsLysGlyGluThrAsn	69
Dd	67	GTTCCGTTTGGTGGAACTGTGACAAAGTCTTGCTGAAGTACAAAAAGGTGAACAAT	126
QY	70	AspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspileLysAspAspGlnIlelle	89
Dd	127	GACTTGTAGTTGTTGAAGACCAGCTGTTAGATCAGACATAAAGTAGTACCAGATCATC	186
QY	90	AsnTrpLeuLeuGluPheArgSerSeriMetTyrlLeuThryLysAspPheGluGlnLeu	109
Dd	187	AACTGGCTGCTAGAATTCGGTTCTTCTGTCATGTACTTGACAAAGACTTTTGAGCACTT	246
QY	110	IleSerIlelleIleuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrr	129
Dd	247	ATCAGTATTATA-----	258
QY	130	LeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSer	149
Dd	258	-----	258
QY	150	MetIleAlaSerHisPheValProArgValIlelleLysGluGlyAspValAspVal	169
Dd	258	-----	258
QY	170	SerAspSerAspAspGluAspAsnLeuProAlaAsnPheAspThrCysHisArgAla	189
Dd	258	-----	258
QY	190	LeuGlnIlelleAlaArgTyrValProSerThrProTrpPheLeuMetProIleLeuVal	209
Dd	258	-----	258

XX WO200172777-A2.  
 XX PD 04-OCT-2001.  
 XX PF 13-MAR-2001; 2001WO-US008117.  
 XX PR 13-MAR-2000; 2000US-0189986P.  
 XX PA (INCY-) INCYTE GENOMICS INC.  
 XX PI Hillman JL, Baughn MR, Yue H, Lal P, Lu DAM, Patterson C;  
 PI Azimzai Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;  
 PI Reddy R;  
 XX WI 2001-570896/64.  
 XX DR P-PSDB; ABB50214.  
 XX PT Novel transcription factor polypeptides, used to treat diseases  
 PT associated with altered activity and expression of TRFX, and to screen  
 PT for agents capable of modulating its activity.  
 XX PS Claim 11; Page 299; 327pp; English.  
 XX CC The present sequence is the coding sequence for a human transcription  
 CC factor. The transcription factor and its coding sequence are useful in  
 CC the diagnosis, treatment and prevention of diseases associated with  
 CC altered expression of the transcription factor e.g. cell proliferative,  
 CC autoimmune/inflammatory, neurological and developmental disorders. A  
 CC number of specific disorders/diseases are given in the specification,  
 CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,  
 CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic  
 CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,  
 CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,  
 CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative  
 CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's  
 CC disease, stroke, and viral, bacterial, fungal and protozoal infections  
 XX SQ Sequence 1461 BP; 413 A; 335 C; 298 G; 415 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: Length: 1461  
 Score: 1124.00 Matches: 256  
 Percent Similarity: 55.44% Conservative: 14  
 Best Local Similarity: 52.57% Mismatches: 52  
 Query Match: 33.03% Indels: 166  
 DB: 5 Gaps: 10

US-09-932-678-2 (1-651) x ABA83038 (1-1461)

QY 195 ArgTyrValProSerThrProTyrPheLeuMetProIleLeuValGluLysPheProPhe 214  
 Db 263 AAATACAAATATAGGCACCGTGGTTCTCATGCCAATACTGGTGGAATAATTTCCATTT 322  
 QY 215 ValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgLysSerVal 234  
 Db 323 GTTCGAAATACAGAGAGACACTGGAATGTACGTTCAATACTTACTAAGGATTAGTGA 382  
 QY 235 TyrPheProThrLeuArgHisGluIleLeuGluLeuIleLeuLysLeuLysLeu 254  
 Db 383 TATTTTCCAACTTGAGCGATGAATTTCTGGAGCTTATTATTGAAAACTACTCAAGCTG 442  
 QY 255 AspValAsnAlaSerArgGlnGlyIleGluAspAlaGluThrAlaThrGlnThrCys 274  
 Db 443 GATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAGCAATCAAACTTGT 502  
 QY 275 GlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluThrGluHis 294  
 Db 503 GTGTGGACAGATTCACGCGGAAGGATTGTTTAATATG----- 538  
 QY 295 GluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGluArgLeu 314  
 Db 538 ----- 538

QY 315 AspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValaspGlyLys 334  
 Db 538 ----- 538  
 QY 335 ValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheaspLysLeu 354  
 Db 538 ----- 538  
 QY 355 LeuLeuProThrHisAlaSerCysHisValGlnPheMetPheTyrLeuCysSerPhe 374  
 Db 538 ----- 538  
 QY 375 LysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTyrLysLeuGlnAspProSer 394  
 Db 539 -----GGATTCGACAGAGCATTTTGGAACTATCTTTGAAAACACTTGCAGGATCCAAGT 592  
 QY 395 AsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAla 414  
 Db 593 AATCCTGCCATCATCAGGCAGGCTGCTGGAATATATATTGGAAGCTTTTGGCAAGAGCT 652  
 QY 415 LysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHis 434  
 Db 653 AAATTTATTCTCTTATTACTCTAAACCATCGCTAGATCTTTTGGTTAACTGGCTGCAC 712  
 QY 435 IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGly 454  
 Db 713 ATATACCTTAATAACAGGATTCGGAAACAAAGGCAATCTGCGATGTTCTCTCCATGGA 772  
 QY 455 ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474  
 Db 773 CCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTATAGACACAAGCAGCTT 832  
 QY 475 LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIle 494  
 Db 833 TTGAGCGAAACCTCAAGAGAGGTTTGCAAGTTCCTCAGAGTCTGAAATTTGAGCGGATA 892  
 QY 495 ValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValIValAsnPheAla 514  
 Db 893 GTGATGAGCCAGCTAAATCCCTGAAGATTTGCCCTCCCTCAGTGTGTTTAACTTTTGTCT 952  
 QY 515 AlaIleThrAsnLysTyr-GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnAr 534  
 Db 953 GCAATCACA---AAGATGAAGACTTTGTG----- 977  
 QY 534 gGlnMetLeuProValIleArgSerThrAlaGlyLysAspSerValGlnIleCysThrAs 554  
 Db 978 -----GATATGGATGGTGGTGATGGT-----TGCAAC 1006  
 QY 554 nProLeuAspThrPhePheProPheAspProCysValLeuLys----- 568  
 Db 1007 AAT-ATCAATTTATTTTATACCATGACCGTGCACTTCAMATGGTTAAGATGTAAGT 1065  
 QY 569 -----ArgSerLysLysPheIleAspPr 576  
 Db 1066 TTTATGTTGTGATTTTACCATAATAAAAAAATTTAGAGGAAAAACAAGTGCCTCC 1125  
 QY 576 oIleTyrGlnValTrpGlu-----AspMetSerAlaGluGl 598  
 Db 1126 ACTTTTG---ATATGGGACTGTAAACATCTCCACCTTCGCTCCCTTCTGCCCCACA 1182  
 QY 588 uLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluAspAspPh 608  
 Db 1183 TCTGGCAAGCTAAGAAAGCCT----- 1204  
 QY 608 eLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrProSer----- 625  
 Db 1205 -----GCTGCTCTCTCTCTGCCACAGCTGGAATTCATACCAACAGCCCTA 1254  
 QY 626 ----SerPheAspThrHisPheArgSerProSerSerValGlySerProProValLe 644  
 Db 1255 GCCTTCCCAACGACCCACATTTTCATCCCATCCCATCGATCCCATCC----- 1306

```
QY 644 uTyrMetGlnProSerPro 650
Db 1307 -----CATCCCATCCC 1318

RESULT 13
AAS92255
ID AAS92255 standard; cDNA; 2410 BP.
XX
AC AAS92255;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #28059.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
(PHYSE-) HYSEQ INC.
XX
PA Drmanac RT, Liu C, Tang YT;
XX
PI
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG28068.
XX

New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity.

Claim 1; SEQ ID NO 28059; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
and in recombinant production of (II). The polynucleotides are also used
in diagnostics as expressed sequence tags for identifying expressed
genes. (I) is useful in gene therapy techniques to restore normal
activity of (II) or to treat disease states involving (II). (II) is
useful for generating antibodies against it, detecting or quantitating a
polypeptide in tissue, as molecular weight markers and as a food
supplement. (II) and its binding partners are useful in medical imaging
of sites expressing (II). (I) and (II) are useful for treating disorders
involving aberrant protein expression or biological activity. The
polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
coding sequences of the invention. Note: The sequence data for this
patent did not appear in the printed specification, but was obtained in
an electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 2410 BP; 672 A; 437 C; 547 G; 754 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.89e-105 Length: 2410
Score: 1074.00 Matches: 220
Percent Similarity: 95.26% Conservative: 1
Best Local Similarity: 94.83% Mismatches: 4
Query Match: 31.56% Indels: 7
DB: 5 Gaps: 0
```

US-09-932-678-2 (1-651) x AAS92255 (1-2410)

QY 427 AspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrIysAla 446  
Db 1 GATCTTTTGGTTAACTGGCTGCACATATACCTTAATACACAGATTTCGGGAACAAGGCA 60

QY 447 PheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThr 466  
Db 61 TTCTGGGATGTTGCTCTCCATGGACCAATTTACTCAGCTGCCAAGCTGTGTTCACACC 120

QY 467 PheValPheArgHisIysGlnLeuLeuSerGlyAsnLeuIysGluGlyLeuGlnIleLeu 486  
Db 121 TTTGTTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAGGTTTCAGTATCTT 180

QY 487 GlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuIysIleCys-Ile 506  
Db 181 CAGAGTCTGAATTTTGGTGGATAGTGATGAGCCAGCTAAATCCCTCTGAAGATATGGCTT 240

QY 506 uProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTy 526  
Db 241 GCCCTCAGTGGTTAACTTTTGTGCTGCAATCAAAATAAGTACAGCTCGTCTTCTGCTA 300

QY 526 rThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAlaGlyGI 546  
Db 301 CACCATCATCGAGAGGAACAATCGCCAGATGCTGCCAGTCTCATTTAGGAGTACCACTGGAGG 360

QY 546 YAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPhe-AspProCysV 566  
Db 361 AGACTCAGTGCAGACCTGCACAAACCCAGTGCACACCTTCTCCCTTTGGATCCCTGTG 420

QY 566 alLeuLysArgSerIysLysPhe-IleAspProIle-Tyr-GlnValTrp-GluAspMet 584  
Db 421 TGCTGAAGAGGTCAAGAAATTCATTGATTCCTATTTATCCAGTATGGGGAACACATG 480

QY 585 SerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGlu 604  
Db 481 AGTGTGAAGAGCTACAGGAGTTCAGAAACCCATGAAAGACATAGTGAAGATGAA 540

QY 605 AspAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrPro 624  
Db 541 GATGATGACTTCTGAAAGGCGAAGTGCACCAGATGATACCGTGATTTGGGATCACCA 600

QY 625 SerSerPheAspThrHisPheArgSerProSerSerValGlySerProProValLeu 644  
Db 601 AGTCTCTTTGACACGCAATTCGGAAGTCTTCAAGTAGTGTGGGCTCCCCACCCCGTGTG 660

QY 645 Tyr-MetGlnProSerProLeu 651  
Db 661 TACACTGCACCCAGTCCCTC 682

RESULT 14  
AAS64829  
ID AAS64829 standard; DNA; 2493 BP.  
XX  
AC AAS64829;  
XX  
DT 15-NOV-2002 (first entry)  
XX  
DE Human tumour suppressor gene #3.  
XX  
KW Human; cancer; gene; ds; tumour suppressor gene; breast cancer; lymphoma;  
KW adenocarcinoma; leukaemia; melanoma; sarcoma; teratocarcinoma;  
KW prostate; cervix; liver; ovary; adrenal gland; heart; lung; colon;  
KW placenta; skeletal muscle; synovial membrane; tonsil; kidney; uterus;  
KW skin; cytostatic.  
XX  
OS Homo sapiens.  
XX  
PN WO200264775-A1.  
XX  
PD 22-AUG-2002.  
XX



```
QY 374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTyrLysLysLeuGlnAspPro 393
Db :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
703 TATCGCACGGGATTCGACAGGCAATTTTGGACATCTTTGGAAAAAATTGCAGGATCCA 762
QY 394 SerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413
Db :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
763 AGTAATCCTCGCATCATCAGGAGGCTGCTGGAAATATATATGGAAAGCTTTTGGCAAGA 822
QY 414 AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTyrLeu 433
Db :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
823 GCTAAATTTATTCCTCTTATTACTGTAAACCATGCTAGATCTTTTGGTTAACTGGCTG 882
QY 434 HisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHis 453
Db :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
883 CACATATACCTTTAATAACAGGATTCGGGAACAAAGGCATTCTGCAATGTTGCTCTCCAT 942
QY 454 GlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln 473
Db :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
943 GGACCATTTTACTCAGCCTGCCAAGCTGTGTCTACACCTTTGTTTTTAGACACAGCAG 1002
QY 474 LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArg 493
Db :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1003 CTTTGGAGCGAAACCTGAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGG 1062
QY 494 IleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePhe 513
Db :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1063 ATAGTGATGAGCCAGCTAAATCCCTGAAGATTGCTGCCCTCAGTGGTTAACTTTTTT 1122
QY 514 AlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn 533
Db :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1123 GCTGCAATCACAAATAAGTACCAGTCGTCTTCTGCTACACCATCATCGAGAGGAACAA 1182
QY 534 ArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThr 553
Db :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1183 CGCCAGATGCTGCCAGTCATTAGGAGTACCCTGGAGGAGACTCAGTGCAGACCTGCACA 1242
QY 554 AsnProLeuAspThrPhePheProPheAspProCysValLeuLysArg 569
Db :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1243 AACCCACTGGACACCTCGTCCCTTTGATCCCTGTGTGCTGAGAGG 1290
```

Search completed: July 18, 2004, 08:57:28

Job time : 701 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.  
OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: July 18, 2004, 06:25:37 ; Search time 6976 Seconds  
(without alignments)  
4044.770 Million cell updates/sec  
Title: US-09-932-678-2  
Perfect score: 3403  
Sequence: 1 MAAPLHLRLPGDAAASSA.....RSPSSSVGSPFVLVMPSP 651

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2 1/USPTO.spool/US09932678/runat 15072004 103603 5206/app query fasta\_1.839  
-DB=GenEmbl -QPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09932678 @CIGN 1 1 4545 @runat 15072004 103603 5206 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUEURY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl :  
1: gb ba : \*  
2: gb htg : \*  
3: gb in : \*  
4: gb om : \*  
5: gb ov : \*  
6: gb pat : \*  
7: gb ph : \*  
8: gb pl : \*  
9: gb pr : \*  
10: gb ro : \*  
11: gb sts : \*  
12: gb sy : \*  
13: gb un : \*  
14: gb vi : \*  
15: em ba : \*  
16: em fun : \*  
17: em hum : \*  
18: em in : \*  
19: em ru : \*  
20: em om : \*  
21: em or : \*  
22: em ov : \*  
23: em pat : \*  
24: em ph : \*  
25: em pl : \*  
26: em ro : \*  
27: em sts : \*  
28: em un : \*

29: em vi : \*  
30: em htg hum : \*  
31: em htg inv : \*  
32: em htg other : \*  
33: em htg mus : \*  
34: em htg pln : \*  
35: em htg rod : \*  
36: em htg mam : \*  
37: em htg vrt : \*  
38: em sy : \*  
39: em htgo hum : \*  
40: em htgo mus : \*  
41: em htgo other : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3403	100.0	2068	9	AF227156	Homo sapi
2	3398	99.9	3756	9	HS272050	Homo sapi
3	3391	99.6	2040	6	AX036047	Sequence
4	3391	99.6	3709	9	BC036182	Homo sapi
5	2372.5	69.7	3166	10	BC034110	Mus muscu
6	2039	59.9	135873	2	AC012547	Homo sapi
7	2039	59.9	192815	9	AC013444	Homo sapi
8	1973	58.0	1423	6	AX427479	Sequence
9	1124	33.0	1461	6	AX274907	Sequence
10	1003	29.5	2627	9	BC006441	Homo sapi
11	1000	29.4	2493	6	AX713860	Sequence
12	1000	29.4	2493	9	AK055742	Homo sapi
13	995.5	29.3	2373	3	AK114871	Ciona int
14	991.5	29.1	2243	10	BC055781	Mus muscu
15	986	29.0	3645	9	AK126166	Homo sapi
16	924	27.2	1168	9	BC009198	Homo sapi
17	783	23.0	2112	3	AY061123	Drosophil
18	712	20.9	437	6	BD073832	5'EST of
19	554.5	16.3	37621	8	SPAC18G6	Z68198 S.pombe chr
20	505	14.8	341	6	BD072397	Secreted
21	499.5	14.7	1671	6	AX489470	Sequence
22	476.5	14.0	2165	8	AY096642	Arabidops
23	474	13.9	1873	8	AY142672	Arabidops
24	456.5	13.4	1815	6	AX507509	Sequence
25	450.5	13.2	2354	8	AK101144	Oryza sat
26	424.5	12.5	2672	8	SCRN3GEN	AK101144
27	424.5	12.5	2843	8	SCYKL125W	Z71927 S.cerevisia
28	424.5	12.5	75335	8	EX842627	Z28125 S.cerevisia
29	420.5	12.4	1884	6	AX596826	Sequence
30	420.5	12.4	1884	6	AX821254	Sequence
31	420.5	12.4	1884	6	AX832284	Sequence
32	403	11.8	191598	2	AC144362	Gorilla g
33	403	11.8	221030	2	AC145403	Gorilla g
34	398	11.7	131839	9	AC112166	Homo sapi
35	398	11.7	157631	2	AC142079	Homo sapi
36	398	11.7	170613	9	AC106788	Homo sapi
37	398	11.7	175691	2	AC007615	Homo sapi
38	398	11.7	185887	2	AC133553	Homo sapi
39	398	11.7	207997	9	AC009093	Homo sapi
40	393.5	11.6	183586	9	AC092562	Papio ham
41	392	11.5	129473	2	AC040158	Homo sapi
42	392	11.5	143783	9	AC139256	Homo sapi
43	392	11.5	184471	9	AC136026	Homo sapi
44	392	11.5	184891	9	AC092137	Homo sapi
45	392	11.5	189770	2	AC139281	Homo sapi

ALIGNMENTS

AF227156	2068 bp	linear	PRI 29-APR-2000
LOCUS			
DEFINITION	Homo sapiens RIN3 mRNA, complete cds.		
ACCESSION	AF227156		
VERSION	AF227156.1	GI:7670099	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE			
AUTHORS	Moorefield,B., Greene,E.A. and Reeder,R.H.		
TITLE	RNA polymerase I transcription factor RIN3 is functionally conserved between yeast and human		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4724-4729 (2000)		
MEDLINE	20243763		
PUBMED	10758157		
REFERENCE			
AUTHORS	Moorefield,B., Greene,E.A. and Reeder,R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-JAN-2000) Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., Seattle, WA 98109, USA		
FEATURES			
source	Location/Qualifiers		
	1..2068		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16p12"		
	69..2024		
	/note="similar to yeast RNA polymerase I transcription factor RIN3"		
	/codon_start=1		
	/product="RIN3"		
	/protein_id="AAF66160.1"		
	/db_xref="GI:7670100"		
	/translation="MAAPLLHTRLPDGAASSAVKVLGASRTGISNRALRENDPFNS PRKTVRFGTVEVLLKYKGETNDFELLKQLLDPDIDKDDIINWLEPRSSIMYL TKDFBQLSIIURLPWLNRQTVVEYLAFGLNLSAQTFLRPLCLSMTASHFVPRV LIKEGDVDSDDEDDNLPAFDTCHRALQIARIYVSTFWFLPILVEKFPFVRKS ERTLSYVHNLRISVYPTFLRHEILIEKLKLDVNASRQIEDAEETATQTCGG DSTSGLENDEDETHETKAGPERLDQMHVAERLDILMSLVLSYMKQVYVDGK VDNKTKLDRLINI FDKLLPTHASCHVQFMFLYCSFKLGFAEAELEHLWKLDQ PNPALIRQAAGNYIGSFLARAKFTPLITKSLCLDLVNWHLHIVLNODSGTKFCDV ALHGFPSYCAQVFTFVFRHKLLSGNLKEGLQSLNPERIVMSQLNPKICLPS VVNFPAATNKQVLFVCYTI IERNRQMLPVIERSTAGDSVQICINPLDTPFPDPCV LKSKKEIDPIYQWEDMSAELEQEFKPKMKDI VEDEDDDFLKGEVFQNDIVIGITP SSFDHFRSPSSVSGPPVLYMQPSPL"		
ORIGIN			
Alignment Scores:			
Pred. No.:	2,05e-316	Length:	2068
Score:	3403.00	Matches:	651
Percent Similarity:	100.00%	Conservative:	0
Best local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0
US-09-932-678-2 (1-651) x AF227156 (1-2068)			
QY	1	MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAla	20
Db	69	ATGGCGGCACCGCTGTTACACACGGCTTTCGGCGGAGATGCGCGCTTCGCTCTGCA	128
QY	21	VallLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp	40
Db	129	GTTAAGAGAGCTGGCGCGTTCGAGGACTGGGATTTCAATATACGTCGATTAGAGATGAC	188
QY	41	PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu	60
Db	189	TTTTTCAATTCTCCCCCAAGAAACTGTTTCGTTTGGTGGACTGTGCACAGAGTCTTG	248
QY	61	LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLysAsnGlnLeuLeuAsp	80

Db	249	CTGAGTACAAAGAGGTGAACAATGACTTTGAGTTGTTGAGAACACGACTGTTAGAT	308
QY	81	ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet	100
Db	309	CCAGACATAAAGGATGACCCAGATCATCAACTGGCTGTAGAAATTCGGTCTTCTATCATG	368
QY	101	TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsn	120
Db	369	TACTTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCTTGGTTGAAT	428
QY	121	ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln	140
Db	429	AGAAGTCAAAACAGTAGTAGGAAGATATTGGCTTTTCTTGGTAAUCTTGTATCAGACAG	488
QY	141	ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal	160
Db	489	ACTGTTTCTCCTCAGACCGGTCTCAGCATGATTGCTTCCCATTTTGTGCTCCCGAGTG	548
QY	161	IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeuPro	180
Db	549	ATCATTAAGGAAGCGGATGATGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCT	608
QY	181	AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr	200
Db	609	GCAAAATTTGACACATGTCACAGAGCTTGCAATAATAGCAAGATATGTACCATCGACA	668
QY	201	ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg	220
Db	669	CGGTGGTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTCGAAAAATCAGAGAGA	728
QY	221	ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg	240
Db	729	ACACTGGAATGTTTACGTTACATAACTTACTAAGGATAGTGTATATTTTCCAACTTGAGG	788
QY	241	HisGluIleLeuGluLeuIleLeuLysLeuLeuLysLeuAspValAsnAlaSerArg	260
Db	789	CATGAANTCTGGAGCTTATATTATGAAAACTACTCAAGTTGGATGTGAATGCATCCCG	848
QY	261	GlnGlyIleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr	280
Db	849	CAGGGTATTGAAGATGCTGAGAAACAGCAACTCAAACTTGTGTGGGACAGATCCACG	908
QY	281	GluGlyLeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyPro	300
Db	909	GAAGATTGTTTAAATATGATGAAGTGAAGAACTGAACATGAACAAAGAGCTGGTCTCT	968
QY	301	GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu	320
Db	969	GAACGGCTCGACCATGCTGATGCTGAGCGGCGCTGGACATCTCTGATGCTTTG	1028
QY	321	ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr	340
Db	1029	GTTTTGTCTCTACATGAAGGATGCTGCTATGTAGATGGTAAAGTTGATAACGCAAAA	1088
QY	341	LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuProThrHisAla	360
Db	1089	AAGGATCTATATCGGACCTGATAAACATCTTTTGACAACTCCTGTTGCCACCCATGCC	1148
QY	361	SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu	380
Db	1149	TCCTGCCATGTACAGTTTTCATGTTTACCTCTCTAGTTTCAANTGGGATTCGACAG	1208
QY	381	AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg	400
Db	1209	GCATTTTGGAACTCTCTGGAAAAAATTCAGGAGCCCAAGTAATCTCTGCCATCATCAG	1268
QY	401	GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle	420
Db	1269	CAGGCTGCTGGAATATATATGGAAGCTTTTGGCAGAGCTAAATTTATTTCTCTATT	1328
QY	421	ThrValLysSerCysLeuAspLeuValAsnTrpLeuHisIleTyrIleAsnAsnGln	440



Db	1329	ACTGTAAATCATGCTAGATCTTTTGGTTAACTGGCTGCCATATACCTTAATACGAC	1388	source	1..3756	/organism="Homo sapiens"
QY	441	AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys	460	gene	/mol_type="mRNA"	/db_xref="taxon:9606"
Db	1389	GATTCGGGAACAAGGCATTCTGCGATGTGCTCTCCATGGACCACTTTTACTCAGCGTCC	1448	CDS	1..3756	/gene="TIF-IA"
QY	461	GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys	480		23..1978	/gene="TIF-IA"
Db	1449	CAAGCTGTGTTCTACACCTTTGTTTTAGACAACAAGCAGCTTTTGGCGGAAACCTGAA	1508		/function="transcription initiation factor"	/evidence=experimental
QY	481	GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn	500		/product="transcription initiation factor IA protein"	/protein_id="CAC07955.1"
Db	1509	GAAGCTTTCAGTATCTTCAGAGTCTGAATTTTGAGCGATAGTAGTGACCGAGTAAAT	1568		/db_xref="GI:10046714"	/db_xref="GOA:Q9H4F0"
QY	501	ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAilaileThrAsnLysTyr	520		/db_xref="SPTREMBL:Q9H4F0"	/translation="MAAPLLHTRLPGDAAASSAVKVLGASRTGISNRALENDFFNS
Db	1569	CCCCTGAAGATTGCTGCTGCTCAGTGGTTAACTTTTGTGCAATCACAAATAAGTAC	1628		PPRKTVRGGVTEVLLIKKGETNDFELKKNQLDPPDKDDQIIINWLLLEFKSSIMYL	TKDFEQLISIIILRPLNRSQTVVEYLAFNLGNLSVAQTFLRPLCLSMIAFSHFVPRV
QY	521	GlnLeuValPheCysTyrThrIleileGluArgAsnAsnArgGlnMetLeuProValIle	540		IIKEGDVDSDDDDNLPANFTCHRALQIIARIYVPSFTWFLMPLIIEVEATQYVRS	ERTLCYVHNLIRISVVPPTLHEILEIIIEKLKLDVNASRQGLIEAEATQYVRS
Db	1629	CAGCTCGTCTTCTGCTACACATCATTCAGAGAAACAATCGCCAGATGCTGCCATCAT	1688		TDSTEGLPNDEDETHETKAGPERLDQVHPVAERLDIILMSLVLSYMKDVQVYDGK	VDNGTKRDLRYDLINIDKLLLPHTASCHVQFPMFYLCFKLGFABAEHLHKKLQD
QY	541	ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe	560		PSNFAILRQDAAGNYTGSFLAKAFIPLITVSKCLDLVNLHLIYLNNDQDSTKAFCDV	ALHGFYSACQAVFYFVRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS
Db	1689	AGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCCTGGACACCTTCTC	1748		VVNFFAATITNKYQVFCYTIITERNRQMLPVIRSTAGDSVQICTNPLDTFPFDPQV	LKRKKFTDPIYQVWEDMSAELOEFKKPKMKDIVEDEDDDDFLKGEVQNDTVIGITP
QY	561	ProPheAspProCysValLeuLysArgSerLysPheIleAspProIleTyrGlnVal	580	3'UTR	1979..3756	/gene="TIF-IA"
Db	1749	CCCTTTGATCCCTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTATCAGGTA	1808	ORIGIN	/evidence=experimental	
QY	581	TrpGluAspMetSerAlaGluLeuGlnGluPheLysLysProMetLysLysAspIle	600	Alignment Scores:		
Db	1809	TGGGAAGACATGAGTGTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAGACATA	1868	Pred. No.:	1.37e-315	Length: 3756
QY	601	ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle	620	Score:	3398.00	Matches: 650
Db	1869	GTGGAAGATGAAGATGATGACTTCTTGAAGCGAGTGCACCAAGATGATACCGTGATT	1928	Percent Similarity:	99.85%	Conservative: 0
QY	621	GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerValGlySer	640	Best Local Similarity:	99.85%	Mismatches: 1
Db	1929	GGGATCACACCAAGCTCCTTTGACAGCATTTCCGAGTCTTCAAGTAGTGTGGCTCC	1988	Query Match:	99.85%	Indels: 0
QY	641	ProProValIleTyrMetGlnProSerProLeu 651		DB:	9	Gaps: 0
Db	1989	CCACCCGTGTGTATCATGCAACCCAGTCCCTC 2021		US-09-932-678-2 (1-651) x HSA272050 (1-3756)		
RESULT 2						
HSA272050						
LOCUS						
DEFINITION	Homo sapiens mRNA for transcription initiation factor IA protein					
ACCESSION	AJ272050					
VERSION	AJ272050.1					
KEYWORDS	TIF-IA; transcription initiation factor.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
TITLE	Bodem,J., Dobrevan,G., Hoffmann-Rohrer,U., Iben,S., Zentgraf,H.,					
JOURNAL	Delius,H., Vingron,M. and Grummt,I.					
MEDLINE	TIF-IA, the factor mediating growth-dependent control of ribosomal					
PUBMED	RNA synthesis, is the mammalian homolog of yeast Rn3p					
REFERENCE	EMBO Rep. 1 (2), 171-175 (2000)					
AUTHORS	11265758					
TITLE	2 (bases 1 to 3756)					
JOURNAL	Bodem,J.J.					
MEDLINE	Direct Submission					
PUBMED	Submitted (07-FEB-2000) Bodem J.J., Molekulare Biologie der Zelle					
REFERENCE	II, Deutsches Krebsforschungszentrum, INF.280, 69120 Heidelberg,					
AUTHORS	GERMANY					
FEATURES	Location/Qualifiers					

Db 443 ACTGTTTTCTCAGACCGGTGCTCAGCATGATTGCTTCCCATTTTGTGCTCCCGAGTG 502  
QY 161 lIeIlleLygluGlyAspValSerAspSerAspGluAspAspValSerAspValSer 180  
Db 503 ATCATTAAGGAGCGCATGTAGATTGTTTCAGATTCTGATGATGAAGATATAATCTTCT 562  
QY 181 AlAAsnPheAspThrCysHISArgAlaLeuGlnIlleAlaAargTyrValProSerThr 200  
Db 563 GCAAAATTTTGACACATGTACAGAGCCTTGCAAAATAATAGCAAGATGATGACCATGACA 622  
QY 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220  
Db 623 CCGTGGTTTCTCATGCCAATACTGCTGGGAAAAATTTCCATTTGTTGAAAAATCAGAGAGA 682  
QY 221 ThrLeuGluCysTyrValHisLenLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
Db 683 ACATCGAAATGTTTACGTTTCATAACTTACTAAGGATTAGTGATATATTTTCCAACTTGAGG 742  
QY 241 HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
Db 743 CATGAAATTTCTGAGAGCTTATATGAAAACTACTCAAGTTGGATGGATGCAATCCCGG 802  
QY 261 GlnGlyIleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
Db 803 CAGGGTATTGAGATGCTGAAGAAAACAGCAACTCAAACTTGTGCTGGGACAGATTCCACG 862  
QY 281 GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300  
Db 863 GAAGGATGTTTAAATATGGAATGAAGATGAAGAACTGAACATGAAACAAAGGCTGTGCTCT 922  
QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
Db 923 GAACGGCTGCACAGATGGTGATCTCTGTAGCGAGCGCTGACATCTGTAGTGTCTTGG 982  
QY 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAenglyLysThr 340  
Db 983 GTTTTGTCTTACATGAAGGATGCTGCTATGTAGATGGTAAGTTGATATACGGCAAAACA 1042  
QY 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla 360  
Db 1043 AAGGATCTATATCGGACCTGATAAATCATCTTTGACAACTCTCTGTGGCCACCCATGCC 1102  
QY 361 SerCysHisValGlnPheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380  
Db 1103 TCTCGCATGTCAGATTTTTCATGTTTACCTCTGTAGTTTCAAAATGGGATTCGAGAG 1162  
QY 381 AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleLeuArg 400  
Db 1163 GCATTTTGTGAACATCTCTGAAAAAATTCAGAGCCCAAGTAAATCTGCGCATCATCAGG 1222  
QY 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420  
Db 1223 CAGGCTGCTGGAATATATATGGAAGCTTTTGGCAAGAGCTAAATTTATCTCTCTTATT 1282  
QY 421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440  
Db 1283 ACTGTAATAATCATGCCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAG 1342  
QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460  
Db 1343 GATTCGGGAAACAAAGGCATTCGCGATGTTGCTCTCATGAGCAATTTTACTCAGCCTGC 1402  
QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
Db 1403 CAAGCTGTGTTCTACRCCTTTTGTGTTTGTAGACACAGAGCTTTTGGCGGAAACCTGAAA 1462  
QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500  
Db 1463 GAAGGTTTGAGTATCTTCAGAGTCTCAATTTTGGCGGATAGTGTAGAGCCAGCTAAAT 1522  
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyr 520  
Db 1523 CCCCTGAAGATTTGCTGCCCTCAGTGTGTTAACTTTTTTGTGCTCAATCAAAATAAGTAC 1582

QY 521 GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetIeuProValIle 540  
Db 1583 CAGCTCGTCTTCTGCTACACCATCATTTGAGAGAAACAATGCGCAGATGCTGCCAGTCATT 1642  
QY 541 ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe 560  
Db 1643 AGGATGACCGCTGGAGGAGCTCAGTGCAGATCTGCACAAACCCGCTGGACACTTCTTC 1702  
QY 561 ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal 580  
Db 1703 CCTTTGATCCCTGCTGCTGAAGAGGTCAAGAAATTCATTGATCTATTTATCAGGTG 1762  
QY 581 TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600  
Db 1763 TGGGAAGATGATGATGCTGAAGAGCTACAGAGCTTCAAGAAACCAATGAAAAAGGACATA 1822  
QY 601 ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle 620  
Db 1823 GTGGAAGATGAAGATGATGATCTTCTGAAAGCGAAGTGCACCAGATGATACCGTGATT 1882  
QY 621 GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerSerValGlySer 640  
Db 1883 GGGATCACCAACCAAGCTCCTTTGACAGCAATTTCCGAAGTCTTCAAGTAGTGTGGGCTCC 1942  
QY 641 ProProValIleuTyrMetGlnProSerProLeu 651  
Db 1943 CCACCCGTGTGTGATGATGCAACCCAGTCCCTC 1975  
RESULT 3  
AX036047 2040 bp DNA linear PAT 15-NOV-2000  
LOCUS Sequence 1 from Patent WO0055316.  
DEFINITION AX036047  
ACCESSION AX036047  
VERSION AX036047.1 GI:11191586  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Grummt, I. and Vingron, M.  
TITLE Rna polymerase i transcription factor tif-ia  
JOURNAL Patent: WO 0055316-A 1 21-SEP-2000;  
DEUTSCHES KREBSFORSCH (DE); GRUMMT INGRID (DE); VINGRON MARTIN (DE)  
FEATURES  
source Location/Qualifiers  
1..2040  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
40..1995  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC16268.1"  
/db\_xref="GI:11191587"  
/translation="MAAPLHTELPGDAASASAVKLGASPTGISNMRALENDFENS  
PRKTVRFGTVEVLKYYKGETNDFELLKQLLDPDKDDQIINWLEFRSIIWYL  
TKDFEQIISIIILRLPWLNSQTVVEYLAPLGNLSAQTVFURPCLSMIASHFVPPRV  
I1KEGDVSDDEDDNLPANFDTCHRALQIARYVPTPWFMPILYVEKFFPVRKS  
ERTLECYVHKLLRSIVYFPTLRHEILELIIKKLLDVNASRQGIEDABETAQTQGG  
TSTEGFNNDDETEHETKAGPERLDQWHPVAERLDILMSLVLSYMKDVCYVQK  
VNGKTKDLYRDLINIFDKLLLPTHASCHVQPMFVLCPSGLFAFAFLEHLWKQLQD  
PSNPALIRAAAGNYGSFLARAKFIPILITVKSCLDLVNWHLIYLNQDSGTAKFCDV  
ALHGPYSACQAVFYXFPFRKQLLSGNLEGLQLQSLNFERIYVNSQLNPKLCILBS  
VNFPAATINQYQVFCYTIILBRNNQKMLPVINRSTAGSDSVQICTNPLDTFFFPDPCV  
LKRKKFIDPIYQVWEDMSAELOQEFKPKMKKDIVEDDDDFLKGVEPQNDTWIGTTP  
SSDFTFRSPSSSVGSPPLYMQPSPL"

## ORIGIN

Alignment Scores: 2.7e-315 Length: 2040  
Pred. No.:

Score:	3391.00	Matches:	649
Percent Similarity:	99.69%	Conservative:	0
Best Local Similarity:	99.69%	Mismatches:	2
Query Match:	99.65%	Indels:	0
DB:	6	Gaps:	0
US-09-932-678-2 (1-651) x AX036047 (1-2040)			
QY	1	MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAspSerSerAla	20
Db	40	ATGGCGGACCGCTGCTTACACAGCGTTTTCGGGAGATGCGCGCTTCCTCTGCA	99
QY	21	ValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp	40
Db	100	GTTAAGAGAGCTGGGCGCGTCGAGGACTGGGATTTCAAAATATATCGTGCATAGAGAAATGAC	159
QY	41	PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu	60
Db	160	TTTTTCAATTCTCCCCCAAGAAAACACTGTTCCGTTTGGTGGAACTGTGACAGAACTCTTG	219
QY	61	LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp	80
Db	220	CTGAAGTCAAAAAGGCTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAGCTGTTAGAT	279
QY	81	ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet	100
Db	280	CCAGACATAAAGGATGACAGATCATCAACTGGCTGCTAGAAATTCGGTTCTTCTATCATG	339
QY	101	TyrLeuThrLysAspPheGluGlnLeuIleSerIleLeuArgLeuProTrpLeuAsn	120
Db	340	TACTTGCACAAAGACTTTGAGCAACTTATCAGTATTATTAAGATTGCCTTGGTTGAAT	399
QY	121	ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln	140
Db	400	AGAACTCAACAGTAGTAGGAGATTTGGCTTTCTTGGTAACTCTGTATCAGACAG	459
QY	141	ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal	160
Db	460	ACTGTTTTTCCTCAGACCGTGCTCAGCATGATGCTGCCATTTTGTGCTCCCGAGTG	519
QY	161	IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAsnLeuPro	180
Db	520	ATCATTAAAGGAAGCGATGTAGATGTTTCAGATTCGTGATGAAGATGATAATCTTCCT	579
QY	181	AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr	200
Db	580	GCAAAATTTGCACATGTCAGAGCCCTTGCAAAATAATAGCAAGATAGTACCATCGACA	639
QY	201	ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg	220
Db	640	CCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTTCGAAAAATCAGAGAGA	699
QY	221	ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg	240
Db	700	ACACTGGATGTAGCTTCATNACTTACTAGGATTAGTATATATTTCCAACTTGAGG	759
QY	241	HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg	260
Db	760	CATGAAATTTCTGGAGCTTATTATTGAAAAAACTACTCAAGTTGGATGTGAATGCATCCCG	819
QY	261	GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr	280
Db	820	CAGGATATTGAAGATGCTGGAAGAAACAGCAACTCAAACTTGTGTGGGACAGATTCCACG	879
QY	281	GluGlyLeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyPro	300
Db	880	GAAGGATTTGTTAATATATGATCAAGATGAAGAACTGAACATGAAACAAAGCGTGGTCT	939
QY	301	GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu	320
Db	940	GAACGGCTCGACAGATGTTGCATCTCTAGCCGAGCGCTGGACATCTCTGATGTCCTTG	999
QY	321	ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr	340

Db	1000	GTTTTGTCTCATCATCAAGGATGTCCTGCTATGTAGATGGTAAGGTTGATAACGGCAAAACA	1059
QY	341	LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla	360
Db	1060	AAGGATCTATATTCGGACCTGATAAACATCTTTTGACAAACCTCTGTGTCGCCACCATGCG	1119
QY	361	SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu	380
Db	1120	TCCTGCCCATGTACAGTTTTCATGTTTACCTCTGTAGTTTCAAAATGGGATTCGACAG	1179
QY	381	AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg	400
Db	1180	GCATTTTGGAAACATCTCTGGAAAAAATTCGAGGACCCCAAGTAATCCTGCCATCATCAG	1239
QY	401	GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle	420
Db	1240	CAGGCTGCTGGAAATTAATTGGAAAGCTTTTGGCAAGAGCTAAATTTATTCTCTTATT	1299
QY	421	ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln	440
Db	1300	ACTGTAAATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATTAACACAG	1359
QY	441	AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys	460
Db	1360	GATTCGGGAACAAAGGCAATTCGCGATGTTGCTCTCCATGGACCACTTTTACTCAGCCTGC	1419
QY	461	GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys	480
Db	1420	CAAGCTGTGTTCTACNCCCTTTGTTTGTAGACACAAAGCAGCTTTTGAGCGGAAACCTGAAA	1479
QY	481	GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn	500
Db	1480	GAAGTTTGGAGTATCTTCAGAGTCTGAATTTTGACCGGATAGTATGATGAGCCAGCTAAAT	1539
QY	501	ProLeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyr	520
Db	1540	CCCCTGAAGATTTGCTGCCCTCAGTGGTTAACTTTTGTGCAATCACAATAAAGTAC	1599
QY	521	GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle	540
Db	1600	CAGCTCGTCTTCTGTCACCATCATTTAGAGGAACAATCGCCAGATGCTGCCAGTCAAT	1659
QY	541	ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe	560
Db	1660	AGGAGTACCGCTGGAGGAGACTCAGTGCAATCTGCACAAACCCGCTGGACACCTTCTTC	1719
QY	561	ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal	580
Db	1720	CCCTTTCATCCCTGTGCTGAAGAGGTCAAAAGAAATTCATTGATCTTATTATCAGGTG	1779
QY	581	TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle	600
Db	1780	TGGGAAGACATGAGTGTGCTGAAGAGCTACAGAGATTCAAGAAACCCATGAAAGAGACATA	1839
QY	601	ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle	620
Db	1840	GTGGAAGATGAAGATGATGACTTCTGAAAGCGGAGTGGCCCCCAGAGATGATACCGTGATT	1899
QY	621	GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerSerValGlySer	640
Db	1900	GGGATCACCAACAGCTCTTTTGACACGCAATTTCCGAAGTCTCTTCAAGTAGTGTGGGTCC	1959
QY	641	ProProValLeuTyrMetGlnProSerProLeu	651
Db	1960	CCACCCGTGTGTACATGCAACCCAGCTCCCTC	1992

RESULT 4  
BC036182  
LOCUS  
DEFINITION Homo sapiens RNA polymerase I transcription factor RN3, mRNA (cDNA clone IMAGE:5298581), partial cds.  
ACCESSION BC036182

BC036182 3709 bp mRNA linear PRI 04-NOV-2003  
clone IMAGE:5298581), partial cds.  
ACCESSION BC036182

VERSION	KEYWORDS	SOURCE	ORGANISM	Query Match:	99.65%	Indels:	0	
				DB:	9	Gaps:	0	
BC036182.1	GI:23243332			US-09-932-678-2	(1-651) x BC036182	(1-3709)		
		Homo sapiens (human)						
		Homo sapiens						
		Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;						
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
REFERENCE		1 (bases 1 to 3709)						
AUTHORS		Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Ustin,I.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,K.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Vallaloon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Binkley,R.W., Touchman,J.W., Green,E.D., Dickinson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smialus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.						
TITLE		Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences						
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)						
MEDLINE		2388257						
PUBMED		12477932						
REFERENCE		2 (bases 1 to 3709)						
AUTHORS		Strausberg,R.						
TITLE		Direct Submission						
JOURNAL		Submitted (31-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA						
REMARK		NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>						
COMMENT		Contact: MGC help desk Email: <a href="mailto:cgaphs-r@mail.nih.gov">cgaphs-r@mail.nih.gov</a> Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN) cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL) cDNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: <a href="http://www-shgc.stanford.edu">http://www-shgc.stanford.edu</a> Contact: (Dickson, Mark) <a href="mailto:mcd@paxil.stanford.edu">mcd@paxil.stanford.edu</a> Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.						
		Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAK Plate: 63 Row: 0 Column: 6 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21361630. Location/Qualifiers 1. .3709 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5298581" /tissue_type="Testis" /clone_lib="NIH MGC_97" /lab_host="DH10B" /note="Vector: pBluescript"						
FEATURES		source						
ORIGIN								
Alignment Scores:								
Pred. No.:		6.29e-315	Length:		3709			
Score:		3391.00	Matches:		649			
Percent Similarity:		99.85%	Conservative:		1			
Percent Local Similarity:		99.69%	Mismatches:		1			

Qy	341	LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla	360
Db	1038	AAGGATCTATATCGCGACTGTATAACATCTTTGACAACTCCTGTGCCACCATGCC	1097
Qy	361	SerCysHisValGlnPheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu	380
Db	1098	TCTGCCATGTACAGTTTTTCAGTTTTTACCTCTCTGTAGTTTCAAATGGATGCCAGAG	1157
Qy	381	AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg	400
Db	1158	GCATTTTTGGAACATCTCTGGAAAAATTTGCAGGACCCCAAGTAATCCTGCCATCATCAGG	1217
Qy	401	GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle	420
Db	1218	CAGGCTGCTGGAATATATTGGAAAGCTTTTTTGGCAAGAGCTAAATTTATTCTCTTATT	1277
Qy	421	ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln	440
Db	1278	ACTGTAAATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACAG	1337
Qy	441	AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys	460
Db	1338	GATTCGGGAACAAAGCATTCGCGATGTGCTCTCCATGGACCATTTTACTCAGCCTGC	1397
Qy	461	GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys	480
Db	1398	CAAGCTGTGTTCTACACCTTTGTTTTTAGACAAAGCAGCTTTTGAGCGGAACCTGAAA	1457
Qy	481	GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn	500
Db	1458	GAAGTTTTCGATATCTTCAGAGTCTGTAATTTTGACGGATAGTAGGACCGCTAAAT	1517
Qy	501	ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr	520
Db	1518	CCCCCTGAAGATTGCTGCCCTCAGTGTGTTAACTTTTTTGCTGCAATCACAAATAAGTAC	1577
Qy	521	GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIle	540
Db	1578	CAGCTCGTCTTCTGTACACCATCATNTGAGAGGAACAATCGCCAGATGTCGCATCATT	1637
Qy	541	ArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhe	560
Db	1638	AGGAGTACCGCTGGAGAGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCCTC	1697
Qy	561	ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVal	580
Db	1698	CCCTTTGATCCCTGTGCTGAAGAGGTCAAAGAAATTCAATTGATCTCTATTATCAGGTG	1757
Qy	581	TrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle	600
Db	1758	TGGGAAGACATGAGTGTGTAAGAGCTACAGGAGTTTCAAGAACCCCATGAAAAGGACATA	1817
Qy	601	ValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIle	620
Db	1818	GTGGAGATGAAGATGATGATCTTCTGAAAGCGAGTGCCCCCAAGATGATACCGTGATT	1877
Qy	621	GlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerValGlySer	640
Db	1878	GGGATCACACCAAGCTCCTTTGACAGCATTTCCGAAAGTCCCTTCAAGTAGTGTGGGCTCC	1937
Qy	641	ProProValLeuTyrMetGlnProSerProLeu	651
Db	1938	CCACCCGTTGTTATCATGCAACCCAGTCCCCCTC	1970
RESULT	5		
LOCUS	BC034110	3166 bp	mRNA linear ROD 16-APR-2003
DEFINITION	Mus musculus expressed sequence AL023001, mRNA (cdna clone IMAGE:4504751), partial cds.		
ACCESSION	BC034110		
VERSION	BC034110.1	GI:21706612	
KEYWORDS	Mus musculus (house mouse)		
SOURCE			

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 3166)

## REFERENCE

## AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, S.I., Max, S.I., Wang, J., Hsieh, P., Dickenson, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Brange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butcherfield, V.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

## REFERENCE

2 (bases 1 to 3166)

Strausberg, R.

Direct Submission

Submitted (02-JUL-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)

Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAC Plate: 44 Row: g Column: 5

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

## FEATURES

## source

1. .3166  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4504751"  
/tissue\_type="Eye, retina, mouse strain C57Bl/6"  
/clone\_lib="NIH MGC\_94"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
<1. .3166  
/gene="AL023001"  
/db\_xref="LocusID:106298"  
/db\_xref="MGI:2146525"  
<1. .1544  
/gene="AL023001"  
/codon\_start=3  
/product="AL023001 protein"  
/protein\_id="AAH34110.1"  
/db\_xref="GI:21706613"

## gene

## CDS

```
/db_xref="LocusID:106298"
/db_xref="MGI:2146525"
translation="PCLSMIAASHFVPPRVIVKEGGIDVSDSDDDDDNLAIPTDCHRA
LQILVYVSTWFLMPLVEXFPPVVKSERLTCEYVHNLRLISLYFPLREILELV
IEKLLDVSVSRQDIEDAEKAAQTCCGGDTTEGLFNMDDEDTDPERKAOEQBNQ
MAHPTABRLDVLICLLLSYIEDVCRVHGKIDNNKLDLYDLISIFDKLLLPHASCH
VQFFPFLCSFKLGFPAEAFLEHMKLQDPNPAITRQAAANYIGSLFARAKFPIIT
VKTCLDLVNLHMYLTNODSGTKAFCDVALGHPFYSAOAVFYVVFVRHKOLLSONL
KQGLVQLSINFERIVLSOLNPLKICLPQVNVNFFAALTNKYQLVFCYTTIMERSOML
PWRSTAGDSVQTCNPLDTPFPDPVCLKSKFELDIYQIWDGSEBELQEPKKS
TKEVVEDEDDDFLKGVEFQSDTVTGLTFSSFDTHRFQSPSSSVGSPPVLYIFQSPPLL
TRIYD"
```

## ORIGIN

```
Alignment Scores:      2,87e-217      Length:      3166
Pred. No.:            2372.50          Matches:     444
Percent Similarity:    93.89%          Conservative: 32
Best Local Similarity: 87.57%          Mismatches:   30
Query Match:          69.72%          Indels:       1
DB:                   10              Gaps:         1
```

US-09-932-678-2 (1-651) x BC034110 (1-3166)

```
QY 146 ProCysLeuSerMetIleAlaSerHisPheValProArgValIleIleLysGluGly 165
Db 3 CCATGTCCTCAGCATGATTGGCTCTCATTTTGTACCTCCCGAGTAATTGTCAGAGGAAGGT 62

QY 166 AspValAspValSerAspSerAspGluAspAspAsnLeuProAlaAsnPheAspThr 185
Db 63 GGCATAGATGTTTCAGATGTTGATGACGAGAGATGACAATCTTCCTGCAATTTTGGACACA 122

QY 186 CysHisArgAlaLeuGlnIleAlaArgTyrValProSerThrProIrpPheLeuMet 205
Db 123 TGTCACAGAGCCTTGCAATAATAACAGATATGTCCCATCGACACCATGGTTCTTAATG 192

QY 206 ProIleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyr 225
Db 183 CCAATACTGGTGAGAAATTCCTGGTTGTGAGGAAGTCCGAGAAACATTTGGAATGTTAT 242

QY 226 ValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGlu 245
Db 243 GTTCATAACTTATAGGATAAGTTATATATTTCCCAACTTTGAGGCGTGAAATTCCTGGAG 302

QY 246 LeuIleIleGluLysLeuLeuLysLeuAspValAlaAsnAlaSerArgGlnGlyIleGluAsp 265
Db 303 CTITGTTATTCGAAAGCTACTCAAGTTAGATGTGAGTGATCGCGGAGGATATAGAGAT 362

QY 266 AlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsn 285
Db 363 GCTGAAGAGAAAGCAGCGCTCAGACTTGTGGTGGACAGATACCAAGAAGACTGTTTAAT 422

QY 286 MetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGln 305
Db 423 ATGGATGAAGATGAGGACACTGACCTCGAAGAAGAGCTGACCCAGGAGCAGCCTAACACC 482

QY 306 MetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMet 325
Db 483 ATGGCTCATCCACCAGCAGCGCTGGAGCTCTCTGTGCTGTGTGCTGTCTCTACAT 542

QY 326 LysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArg 345
Db 543 GAGGATGTCGCGCTGTGACCGGTAAATTTGATTAACAATAAACAAGGATTTTATACCGT 602

QY 346 AspLeuIleAsnIlePheAspLysLeuLeuProThrHisAlaSerCysHisValGln 365
Db 603 GATCTGATATCCATCTTTGCAAACTCTGTTCGCCACACATGCTCCCTCGCATGTACAG 662

QY 366 PhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHis 385
Db 663 TTCTTCATGTTTTCCTCTCGAGCTTCAAGTTGGGATTTGCAAGCATTTTTCGAACAT 722

QY 386 LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleArgGlnAlaAlaGlyAsn 405
```

```
Db 723 CTCGTGAAAAGATTGTCAGGATCCAAATAACCCCGCCATCATCAGGAGGCTGCTGCAAAAT 782
QY 406 TyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys 425
Db 783 TATATTGGGAGCTTTTGGCCAGAGCTAAATTTATTCCTCTTATCATCTGTGAAGACATGC 842
QY 426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLys 445
Db 843 CTGGATCTCTTGGTTAACTTGGCTGCACATGATCCTTACTAACACAGATTCGGGAAACAAG 902
QY 446 AlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyr 465
Db 903 GCTTTTGTGACGTTGCATCTCCATGACCATTTATTTCAGCTGCTGCCAAGCTGTGTTCTAC 962
QY 466 ThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyr 485
Db 963 ACTGTTGTTTTTAGACACAGCAGCTTTTGGTGGAAACTTGAAGCAAGGTCTACAGTAT 1022
QY 486 LeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCys 505
Db 1023 CTTTCAGAGTCTAAATTTTTCAGCGCATTTGTTGAGCCAGCTGAACCCACTGAAGATCTGC 1082
QY 506 LeuProSerValValAsnPhePheAlaIleThrAsnLysTyrGlnLeuValPheCys 525
Db 1083 CTGCCGCAAGTGGTTAAATTTCTTGTGCTATCACAATAAATACCAAGCTGGTGTCTGC 1142
QY 526 TyrThrIleIleGluArgAsnArgGlnMetLeuProValIleArgSerThrAlaGly 545
Db 1143 TACACCATCATCGAGAGGAAACAGTCGCCAGATGCTCCCGTTATTCGCGACACCGCTGGT 1202
QY 546 GlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCys 565
Db 1203 GGAGACTCCGTGCAGACCTGCACCAACCCACTGGACACTTCTTCCCTTTGACCCCTGT 1262
QY 566 ValLeuLysArgSerLysLysPheIleAspProIleTyrGlnValIrpGluAspMetSer 585
Db 1263 GTGCTTAAGAGGTCGAAAGAGTTTCATTGATCTCTATTATTCAGATTTTGGGAAGATGGAGT 1322
QY 586 AlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluAsp 605
Db 1323 GCTGAAGAGCTTCAGAGATTGAAGAAATCTCTAAAGAGAGGAGTAGTGGAGGATAGAGAT 1382
QY 606 AspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrProSer 625
Db 1383 GATGACTTTTGAAGGCGAGGTGCCCGAGAGTGACACAGTGAGTGCCTTACTCCGAGC 1442
QY 626 SerPheAspThrHisPheArgSerProSerSerSerValGlySerProProValLeuTyr 645
Db 1443 TCTTTTGATACCCACTTCCAAAGTCTTCCAGTAGTGTGGCTCCCTCTCTGTGTGTAT 1502
QY 646 MetGln--ProSerProLeu 651
Db 1503 ATACAGGCGCAGTCTCCACTC 1523

RESULT 6
AC012547 135873 bp DNA linear HTG 30-MAR-2000
LOCUS Homo sapiens clone RP11-11K1, WORKING DRAFT SEQUENCE, 14 unordered
DEFINITION pieces.
ACCESSION AC012547
VERSION AC012547.2 GI:7341888
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 135873)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-11K1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 135873)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
```

Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

# TITLE JOURNAL

## COMMENT

Submitted (29-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 30, 2000 this sequence version replaced gi:61391119.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L3178

Center clone name: 11 K 1

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 127596 bases at least Q40

Consensus quality: 130929 bases at least Q30

Consensus quality: 132388 bases at least Q20

Insert size: 133000; agarose-fp

Insert size: 134573; sum-of-contigs

Quality coverage: 4.9 in Q20 bases; agarose-fp

Quality coverage: 4.9 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 14 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 1516: contig of 1516 bp in length  
\* 1517 1616: gap of 100 bp  
\* 1617 3136: contig of 1520 bp in length  
\* 3137 3236: gap of 100 bp  
\* 3237 5148: contig of 1912 bp in length  
\* 5149 5248: gap of 100 bp  
\* 5249 7742: contig of 2494 bp in length  
\* 7743 7842: gap of 100 bp  
\* 7843 10651: contig of 2808 bp in length  
\* 10651 10750: gap of 100 bp  
\* 10751 16194: contig of 5444 bp in length  
\* 16195 16294: gap of 100 bp  
\* 16295 25751: contig of 9457 bp in length  
\* 25752 25852: gap of 100 bp  
\* 25853 37209: contig of 11358 bp in length  
\* 37210 37309: gap of 100 bp  
\* 37310 47828: contig of 10519 bp in length  
\* 47829 47929: gap of 100 bp  
\* 47930 64240: contig of 16311 bp in length  
\* 64240 64340: gap of 100 bp  
\* 64340 77847: contig of 13507 bp in length  
\* 77847 77947: gap of 100 bp  
\* 77947 95311: contig of 17365 bp in length  
\* 95312 95411: gap of 100 bp  
\* 95412 108930: contig of 13519 bp in length

\* 108931 109030: gap of 100 bp  
\* 109031 135873: contig of 26843 bp in length.

## FEATURES

source  
1. 135873  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-11K1"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. 1516  
/note="assembly\_fragment"  
1617..3136  
/note="assembly\_fragment"  
3237..5148  
/note="assembly\_fragment"  
5249..7742  
/note="assembly\_fragment"  
7843..10650  
/note="assembly\_fragment"  
10751..16194  
/note="assembly\_fragment"  
clone end:SP6  
vector side:right  
16295..25751  
/note="assembly\_fragment"  
25852..37209  
/note="assembly\_fragment"  
37310..47828  
/note="assembly\_fragment"  
47929..64239  
/note="assembly\_fragment"  
64340..77846  
/note="assembly\_fragment"  
77947..95311  
/note="assembly\_fragment"  
95412..108930  
/note="assembly\_fragment"  
109031..135873  
/note="assembly\_fragment"

## ORIGIN

Alignment Scores:  
Pred. No.: 5.89e-183 Length: 135873  
Score: 2039.00 Matches: 453  
Percent Similarity: 77.71% Conservative: 56  
Best Local Similarity: 69.16% Mismatches: 120  
Query Match: 59.92% Indels: 32  
DB: 2 Gaps: 13

US-09-932-678-2 (1-651) x AC012547 (1-135873)

QY 2 AlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaLaserSerAlaVal 21  
Db 96743 GCAACCAGGCTGCTTCTCACATGGTTGCCGGGAA-GCAGAGGCTTTCAGTCTCTGGTGC 96801  
QY 22 LysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPhe 41  
Db 96802 AAGAGCTGGGTATGTCATGCACCATGATTTTCAGATATGTTGCATTCAGAAATGATTTT 96861  
QY 42 PheAsnSerProArgLysThrValArgPheGlyGlyThrValThrGluValLeuLeu 61  
Db 96862 TCCAGCTCTCCCCAAGAAATAATGTTTCATCTTGATGGAACGGGACAGAGCTCTGGCG 96921  
QY 62 LysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspPro 81  
Db 96922 AAGACAAAGAGTGAACAAATTTAAGCTTTGAGCTGTTGAAGAGCCATTTGCGATTCA 96981  
QY 82 AspIleLysAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMetTyr 101  
Db 96982 AACATAAAGGATGACCATCTTGAACCTGGTGTGAAATTTTGTCTCTTCTATCATGTAC 97041  
QY 102 LeuThrLysAspPheGluLeuLeuSerIleLeuLeuArgLeuProTrpLeuAsnArg 121



```
Db 97042 TTGGCAAAAGACTTTGAGCAGCTCTCAGTATTACATTAGATGCTGGTTGAATAGA 97101
QY 122 SerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThr 141
Db 97102 ATTCAACAGTGTGGAGAGTAGTATTGGCTTCTCTGGGTAGTCTGTAAACAAC-AGA-- 97157
QY 142 ValPheLeuArgProCysLeuSerMetIleAlaSerHisPheVal---ProProArgVal 160
Db 97158 CTCTCTTAGCCCAAGTCTCAGCATATGTCTTCCATTGTGGCGTACCCTCCTGGAATG 97217
QY 161 IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeuPro 180
Db 97218 ATCATTAAGAGAGCCCATGTAATGTTTCAGATTCTGATGAGAGATGATTATCTCTCT 97277
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleAlaArgValProSerThr 200
Db 97278 GCAGATTTTGACACAGCTCATACAACCTTACAATAACAGCAAGATATGCCACCTGGCA 97337
QY 201 ProTrioPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220
Db 97338 CTGTGTTTCTTTGCCCACACTGCTGGTGGAAACATTTCCATTCTGT---CAATCAGAGAGA 97394
QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240
Db 97395 ACACCTGGAAATTTACATTTCAATTAACCTTACTAAAGATTAGTATATACCTTCCAACTTGAGG 97454
QY 241 HisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuAspValAlaSerArg 260
Db 97455 AATGAATTTCTGAGCTTATTATAAATCTATTCAATGGATGGATGAATACATTCTGG 97514
QY 261 GlnGlyIleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280
Db 97515 CATGATATTCAAGATGCTGAGGAAACAGCAGCTCAAACTTGTGGTGGGACAGATTCATG 97574
QY 281 GluGlyLeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyPro 300
Db 97575 GAATGATTTGTTCATATGACAGAGATGAAGAGACT-----GTTTCTAAGGCTGACTCC 97628
QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320
Db 97629 AATAGCTCGGTCAGATGGTGACCTGTAGCTGAATGCTGGACACCTCGCTGTCTTTG 97688
QY 321 ValLeuSerTyrMetLysAspValCysTyrValAsp-----GlyLysValAspAsnGly 338
Db 97689 CTTTTCAGTACTTGAATGATGCTGCTATGTAAA-ATACATGGTAAGGTGATGAACAAC 97747
QY 339 LysThrLysAspLeu-----TyrArgAspLeuIleAsnIlePheAspLysLeuLeu 356
Db 97748 AAATAAAGAGATTT-ATAATATATCAAACTGATAAACAATATTGACAAA----- 97797
QY 357 ProThrHisAlaSerCysHisValGlnPheMetPheTyrLeuCysSerPheLysLeu 376
Db 97798 ---CCCTT-GCTTCTGTACAGTACAGTTTATGTTTATCTTTTACCTCTGTGGCTTAAATG 97853
QY 377 GlyPheAlaGluAlaPheLeuGluHisLeuTrioPheLysLeuGlnAspProSerAsn 395
Db 97854 CGACTCTCAGAGCATTTTGGAAACGCTCTCTGGAAAAAATAATGCAAGATCCAAATAAT 97913
QY 396 ProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLys 415
Db 97914 CTGTCTATCATCAGTGAGCTGCTGGAAATATTATTGGAAGCTCTTTGGCAAAAGCTAAA 97973
QY 416 PheIleProLeuIle---ThrValLysSerCysLeuAspLeuLeuValAsnTrioPheHis 434
Db 97974 TTCAATTCCTCTCATTAACAACATAFAAAATCATGCTAGGCTTTTCGGTTAACTGGCCAC 98033
QY 435 IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGly 454
Db 98034 AATACCTTACTAACACAGATCTGAGAATGAAGGCTTTTGTGATATTCTCTCCATCAA 98093
QY 455 ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474
Db 98094 ACATTTTACTCAGCCTTCAAGCTGTGTTCCACACCTTGTGTTTATGACACAGCAACTT 98153

QY 475 LeuSerGlyAsnLeuLysGluGlyIleuGlnTyrLeuGlnSerLeuAsnPheGluArgIle 494
Db 98154 TTAAGCTGAACCTGAAGAAGGTTTGAGATGTTCTTCACTGCCTAGATTTTTTAAGGGATA 98213
QY 495 ValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAla 514
Db 98214 GTGACTAGGACAGCTAATCTCTGAATATTGCTGCCCTCAGTGGTTAACTTATTGCT 98273
QY 515 AlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn-Ar 534
Db 98274 GCTGTCCAGCAATAAATATCAGCTAGTC---TGCTACACTATCATTCAGAGGAACAACGGC 98330
QY 534 gGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAs 554
Db 98331 C-----AGAAATACAGCTGGAGAGACTCGGTGTGAACGTGCACCAA 98372
QY 554 nProLeuAspThrPhePheProPheAspProCysValLeuLysArgSerLysLysPheIle 574
Db 98373 CCATTGGACGCTTCTTCCCTTTCATCTTGTGGCTTAAGCAGTCAAGGAAGTCAT 98432
QY 574 eAspProIleTyrGlnValTrioPheGluAspMetSerAlaGluGluLeuGlnGluPheLys 594
Db 98433 TGATTTCTATTATTCAGATATGGAGACATAAAGTGTGGAAGAGCTTCAGAGATTAAAGA 98492
QY 594 sProMetLysLysAspIleValGluAspGluAspAspPheLeuLysGlyGluValPr 614
Db 98493 ACCTATGAAAAGAGAG---ATGGAATAAGATGAGATGATCTTTTGAAGGTGAAGTTCC 98549
QY 614 oGlnAsnAspThrValIleGlyIleThrProSerPheAspThr-----HisPheAr 632
Db 98550 CCCAAATGATACCATGATCGGATATATACCGACCACTTTGATGC-GATGCACATTTCCA 98608
QY 632 gSerProSerSerValGlySerProProValLeuTyrMet 646
Db 98609 AAGTCCTTCAAGTAGTGTGGCTCCCACTGTGTACACCTG 98651

RESULT 7
AC013444 192815 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-470B22 from 2, complete sequence.
DEFINITION AC013444
ACCESSION AC013444
VERSION AC013444.8 GI:18093267
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 192815)
AUTHORS Sulston,J.E. and Waterston,R.
TITLES Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 192815)
AUTHORS Walligorski,J., Haakenson,W. and Spalding,L.
TITLES The sequence of Homo sapiens BAC clone RP11-470B22
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 192815)
AUTHORS Waterston,R.H.
TITLES Direct Submission
JOURNAL Submitted (11-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 192815)
AUTHORS Waterston,R.
TITLES Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 9, 2002 this sequence version replaced gi:13518264.
-----Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
```



Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0470B22  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McEherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletzer de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC024616; the clone sequenced to the right is RP11-190J23. Actual start of this clone is at base position 1 of RP11-470B22; actual end is at base position 192815 of RP11-470B22.

Unresolved homopolymeric run from base position 164727 to 164753.

#### FEATURES

##### source

1..192815  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"

/clone="RP11-470B22"  
 /clone\_lib="RPCI-11"

1..575

/rpt\_family="L1"

611..878

/rpt\_family="ERV1"

879..1032

/rpt\_family="(TAGA)n"

1046..1092

/rpt\_family="ERV1"

1987..2346

/rpt\_family="L2"

2737..2972

/rpt\_family="L1"

3702..3739

/rpt\_family="AT\_rich"

4870..4891

/note="match to EST BE071256 (NID:g8416143)"

6880..6992

/rpt\_family="MIR"

8677..8934

/rpt\_family="Alu"

8773..8800

/rpt\_family="AT\_rich"

repeat\_region 8955..8989  
 /rpt\_family="AT\_rich"  
 repeat\_region 9207..9247  
 /rpt\_family="L2"  
 repeat\_region 9342..9773  
 /rpt\_family="L2"  
 repeat\_region 9977..10011  
 /rpt\_family="(TTTA)n"  
 repeat\_region 10976..11265  
 /rpt\_family="Alu"  
 repeat\_region 12178..12207  
 /rpt\_family="(T)n"  
 repeat\_region 12362..12390  
 /rpt\_family="AT\_rich"  
 repeat\_region 12702..12731  
 /rpt\_family="(TG)n"  
 repeat\_region 12731..12758  
 /rpt\_family="(GA)n"  
 repeat\_region 12758..12785  
 /rpt\_family="(TG)n"  
 repeat\_region 13033..13054  
 /rpt\_family="(CAAAA)n"  
 repeat\_region 13195..13699  
 /rpt\_family="L1"  
 repeat\_region 13543..13567  
 /rpt\_family="AT\_rich"  
 repeat\_region 14142..14305  
 /rpt\_family="MIR"  
 repeat\_region 14500..14534  
 /rpt\_family="AT\_rich"  
 repeat\_region 14617..14735  
 /rpt\_family="MIR"  
 repeat\_region 14736..14836  
 /rpt\_family="MER2\_type"  
 repeat\_region 14837..15106  
 /rpt\_family="Alu"  
 repeat\_region 15107..15317  
 /rpt\_family="MER2\_type"  
 repeat\_region 15318..15385  
 /rpt\_family="MIR"  
 repeat\_region 15578..16066  
 /rpt\_family="ERV1"  
 repeat\_region 16072..16112  
 /rpt\_family="AT\_rich"  
 repeat\_region 18192..18222  
 /rpt\_family="AT\_rich"  
 repeat\_region 18612..18899  
 /rpt\_family="Alu"  
 misc\_feature 19181..19485  
 /note="match to EST R89091 (NID:g953918) Yq02f04.s1"  
 misc\_feature 19879..21714  
 /note="similar to Rattus norvegicus EST Y16973 (NID:g4727035)"  
 misc\_feature 19890..20284  
 /note="match to EST R89197 (NID:g954024) Yq02f04.r1"  
 misc\_feature 19915..20369  
 /note="similar to Homo sapiens EST BG535886 (NID:g13527431)"  
 misc\_feature 19923..20378  
 /note="similar to Homo sapiens EST AA213789 (NID:gi1812416) z191a11.s1"  
 misc\_feature 19923..20369  
 /note="similar to Homo sapiens EST BG428305 (NID:gi1334811)"  
 misc\_feature 20091..20378  
 /note="similar to Homo sapiens EST A1742153 (NID:g5110441) wg49f01.x1"  
 misc\_feature 20158..20664  
 /note="similar to Homo sapiens EST AV703279 (NID:gi10720608)"  
 misc\_feature 20243..21062  
 /note="similar to Mus musculus EST BG295610 (NID:gi13057417)"

[illegible]

```

QY 574 eAspProIleTyrGlnValTrpGluAspMetSerAlaGluGluLeuGlnGluPheLysly 594
Db 21505 TGATTCATTTATCATGATGGAGACATAGTGTGGAAGACTTCAGGAGTTTAAAGAA 21564
QY 594 sProMetLysLysAspIleValGluAspGluAspAspPheLeuLysGlyGluValPr 614
Db 21565 ACCTATGAAAAGGAG--ATGGAAAAGATGGAGATGATCTTTTGAAGGTGAAGTTCC 21621
QY 614 oGlnAsnAspThrValIleGlyIleThrProSerSerPheAspThr-----HisPheAr 632
Db 21622 CCCAATGATACCATCATCGGATTATACCGAGCCACTTTTGATGC-GATGCACATTTCCA 21680
QY 632 gSerProSerSerValGlySerProProValLeuTyrMet 646
Db 21681 AAGTCTCTCAAGTAGTGTGGGTCCCACTGTGTGTACACCTG 21723

RESULT 8
AX427479 1423 bp DNA linear PAT 20-JUN-2002
LOCUS AX427479
DEFINITION Sequence 52 from Patent WO0123558.
ACCESSION AX427479
VERSION AX427479.1 GI:21537625
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Hodgson,D.M., Lincoln,S.E., Russo,F.D., Spiro,P.A., Banville,S.C.,
Bratcher, Shawn,R., Dufour,G.F., Cohen,H.J., Rosen,B.H., Shah,P.,
Chalup,M.S., Hillman,J.L., Jones,A.L., Yu,J.Y., Greenawalt,L.B.,
Panzer,S.R., Roseberry,A.M., Wright,R.J., Chen,W., Liu,T.,
Yap,P.E., Stockdreher,T.K., Amshy,S. and Fong,W.T.
Human secretory molecules
Patent: WO 0123558-A 52 05-APR-2001;
Incyte Genomics, Inc. (US)

FEATURES
Location/Qualifiers
1..1423
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/noife="Incyte ID No: 428745.2.dec"

ORIGIN
Alignment Scores:
Pred. No.: 2,05e-179 Length: 1423
Score: 1973.00 Matches: 393
Percent Similarity: 80.53% Conservative: 0
Best Local Similarity: 80.53% Mismatches: 4
Query Match: 57.98% Indels: 1
DB: 6 Gaps: 1

US-09-932-678-2 (1-651) x AX427479 (1-1423)

QY 31 IleSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
Db 194 ATTTCAATATGCGTGCAATAGAAATGATTTTTCATTTCTCCCAAGAAAACCTGTT 253
QY 51 ArgPheGlyGlyThrValThrGluValLeuLysTyrLysGlyGluThrAsnAsp 70
Db 254 CGGTTTGGTGAACCTGTGACAGAGCTCTGTGAGTACAAAAGGGTGAACAAATGAC 313
QY 71 PheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleLeuAsn 90
Db 314 TTTGAGTTGTTTGAAGAACCACTGTTAGATCAGACATTAAGGATGACCATCATCAAC 373
QY 91 TrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIle 110
Db 374 TGGCTGCTAGAAATTCGGTCTCTCATCATGATGACTTGTGACAAAGACTTTGAGCAACTTATC 433
QY 111 SerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrIleu 130

```

```

Db 434 AGTATTATATTAGATTGCTTTGTTGATTGATAGAAAGTCAAAACAGTAGTGAAGAGTATTG 493
QY 131 AlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
Db 494 GCTTTTCTTGGTAACTCTGTATCAGACAGACTGTTTTCTCAGACCGGTCTCAGCATG 553
QY 151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspValAspValSer 170
Db 554 ATTGCTTCCCACTTTGTGCTCTCCCGAGTGATCAATTAAGGAAGCGATGTAGATGTTTCA 613
QY 171 AspSerAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeu 190
Db 614 GATTCGTGATGATGAAGATGATTAATCTCCTGCAAAATTTTGACACATGTCCAGAGACCTTG 673
QY 191 GlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProIleLeuValGlu 210
Db 674 CAATATAGTACAGATATGTACCATCGACACCGTGTCTCATGCCAATATCTGTTGGA 733
QY 211 LysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeu 230
Db 734 AAATTTCCATTTGTTGAAAATCAGAGAGACACTGGATGTGTAGCTTCATTAACCTACTA 793
QY 231 ArgIleSerValTyrPheProThrIleuArgHisGluIleLeuGluLeuIleLeuLys 250
Db 794 AGGATTAGTGTATATTTTCCAACTTGAGGCATGAAATCTCGAGCTTATTATTGAAAA 853
QY 251 LeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluThrAla 270
Db 854 CTACTCAAGTTGGATGTGAATGTCATCCCGCAGGCTATTTGAAGATGCTGAAAGAACAGCA 913
QY 271 ThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGlu 290
Db 914 AATCAAACTTGTGGTGGACAGATTCCACGGAAGGATGTTTAAATG-----961
QY 291 GluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProVal 310
Db 961 -----961
QY 311 AlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyr 330
Db 961 -----961
QY 331 ValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIle 350
Db 961 -----961
QY 351 PheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyr 370
Db 961 -----961
QY 371 LeuCysSerPheLysLeuGlyPheAlaGluAla-PheLeuGluHisLeuTrpLysLysLe 390
Db 962 -----GGATTGCGAGAGGCCAATTTTGGACATCTTTGGAACAACTT 1003
QY 390 uGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPh 410
Db 1004 GCAGGATCCAAAGTAACTCTGCCATCATCAGCAGCGCTGCTGGAATATATATTGGAAGCTT 1063
QY 410 eLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuVa 430
Db 1064 TTTGGCAAGAGCTAAATTTATTTCTCTTATTACTGTAAACCATGCTCCTAGATCTTTTGGT 1123
QY 430 lAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspVa 450
Db 1124 TAATCTGGCTGCACATATACCTTAATACAGGATTCGGGAACAAAGGATCTTCGCGATG 1183
QY 450 lAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheAr 470
Db 1184 TGCTCTCCATGGACCAATTTTACTCAGCTTCGCAAGCTGTGTTCTACACCTTTGTTTTAG 1243
QY 470 gHisLysGlnIleLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAs 490
Db 1244 ACAACAGCAGCTTTTGAGCGGAACCTGAAAGAGGTTTCAGTATCTTCAGAGCTCTGAA 1303

```

QY 490 nPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValVa 510  
Db 1304 TTTTGAGCGGATAGTATGATGAGCAGCAATAATCCCTGAAGATTTGCCTCCCTCAAGTGGT 1363

QY 510 lAsnPhePheAlaAlaIleThr 517  
Db 1364 TAACTTTTGTGCAATCACA 1385

RESULT 9  
AX274907  
LOCUS AX274907 1461 bp DNA linear PAT 29-OCT-2001  
DEFINITION Sequence 172 from Patent WO0172777.  
ACCESSION AX274907  
VERSION AX274907.1 GI:16547539

KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Hillman, J.L., Baughman, M.R., Yue, H., Lal, P., Lu, D.A., Patterson, C., Azimzai, Y., Bandman, O., Tang, Y.T., Mathur, P., Shah, P., Au-Young, J. and Reddy, R.

TITLE Transcription factors  
JOURNAL Patent: WO 0172777-A 172 04-OCT-2001;  
INCYTE Genomics, Inc. (US)  
FEATURES  
Location/Qualifiers  
1..1461  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 2646274CB1"

## ORIGIN

Alignment Scores:  
Pred. No.: 6.51e-98 Length: 1461  
Score: 1124.00 Matches: 256  
Percent Similarity: 55.44% Conservative: 14  
Best Local Similarity: 52.57% Mismatches: 52  
Query Match: 33.03% Indels: 166  
DB: Gaps: 10

US-09-932-678-2 (1-651) x AX274907 (1-1461)

QY 195 ArgTyrValProSerThrProTyrPheLeuMetProIleLeuValGluLysPheProPhe 214  
Db 263 AAATACAAAATTAGGCGCACCGTGGTTCTCATGCCAATACTGGTGGAAAAATTTCCATTT 322

QY 215 ValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerVal 234  
Db 323 GTTCGAAATCAGAGAACACTGGAAATGTTAGTTTCATTAACCTACTTAAGATTAGTGA 382

QY 235 TyrPheProThrLeuArgHisGluIleLeuGluLeuIleLeuLysLeuLeuLysLeu 254  
Db 383 TATTTTCCAACTTGAGCGCATGAAATCTCGAGCTTATTATGAAAACTACTCAAGCTG 442

QY 255 AspValAsnAlaSerArgGlnGlyIleGluAspAlaGluThrAlaThrGlnThrCys 274  
Db 443 GATGTGAATCATCTCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAATCAAACTGT 502

QY 275 GlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGluThrGluHis 294  
Db 503 GGTGGGACAGATCCACGAGAGGATTTGTTTAATG----- 538

QY 295 GluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGluArgLeu 314  
Db 538 ----- 538

QY 315 AspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLys 334  
Db 538 ----- 538

QY 335 ValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeu 354  
Db 538 ----- 538

QY 355 LeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPhe 374  
Db 538 ----- 538

QY 375 LysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTyrLysLeuGlnAspProSer 394  
Db 539 -----GGATTCCGACAGGCATTTTGGAAACATCTTTGGAATAACTTGCAGGATCCAAGT 592

QY 395 AsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAla 414  
Db 593 AATCTGCCATCATCAGGCAGGCTGCTGGAATTAATATTGGAAGCTTTTGGCAAGAGCT 652

QY 415 LysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTyrLeuHis 434  
Db 653 AAAATTTATTCTCTTATTACTGTAAACCATGCTAGATCTTTTGGTTAACTGGCTGCAC 712

QY 435 IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGly 454  
Db 713 ATATACCTTAAATACAGGATTCGGGAACAAAGCATCTCGAATGTTGCTCTCCATGGA 772

QY 455 ProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474  
Db 773 CCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTATAGACAACAGCAGCTT 832

QY 475 LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIle 494  
Db 833 TTGAGCGGAAACCTGAAAGAGGTTTGCAGTAGTATCTCCAGAGCTCTGAATTTTGAGCGGATA 892

QY 495 ValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAla 514  
Db 893 GTGATGAGCCAGCTAAATCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTGTCT 952

QY 515 AlaIleThrAsnLysTyr-GlnLeuValPheCysTyrThrIleIleGluArgAsnAsnAr 534  
Db 953 GCAATCACA--AAGATGAAGACTTGTG----- 977

QY 534 gGlnMetLeuProValIleArgSerThrAlaGlyLysAspSerValGlnIleCysThrAs 554  
Db 978 -----GATATGGATGGTGGTGATGGT-----TGCACAAC 1006

QY 554 nProLeuAspThrPhePheProPheAspProCysValLeuLys----- 568  
Db 1007 AAT-ATCAATTTATTTATACCACTGAACCGTGCACTTCAAAATGGTTAAGATGGTAAGT 1065

QY 569 -----ArgSerLysLysPheIleAspPr 576  
Db 1066 TTTATGTTGTGTATTTTACCATAATAAAAAAATTTGAGAGGGAACAAAGCTGTGCCTCC 1125

QY 576 olleTyrGlnValTyrGlu-----AspMetSerAlaGluGl 588  
Db 1126 ACTTTTG--ATATGGGACTGTAAACATCTCCACCTTCGCTCTCCCTCTCTGCCCCACA 1182

QY 588 uLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluAspAspPh 608  
Db 1183 TCTGGCAAGCTAAGAAAGCT----- 1204

QY 608 eLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrProSer----- 625  
Db 1205 -----GCTGCTCTCTCTCTGGCACCAAGCTGGAATTCATACCCCAACAGCCCTA 1254

QY 626 ----SerPheAspThrHisPheArgSerProSerSerValGlySerProValLe 644  
Db 1255 GCCTTCCCACGACCCACATTTTCATCCCATCCCATCCCATCCCATCCCATCCCATCC 1306

QY 644 uTyrMetGlnProSerPro 650  
Db 1307 -----CATCCCCATCCC 1318

RESULT 10

BC006441	2627 bp	mRNA	linear	PRI 02-JUL-2003
LOCUS				
DEFINITION	Homo sapiens similar to RNA polymerase I transcription factor RRN3, mRNA (cDNA clone IMAGE:3351791), partial cds.			
ACCESSION	BC006441			
VERSION	1	GT:13623642		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 2627)			
	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Woxley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodrigues, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalski, U., Smalios, D.E., Smerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.			
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)			
MEDLINE	22388257			
PUBMED	12477932			
REFERENCE	2 (bases 1 to 2627)			
AUTHORS	Strausberg, R.			
TITLE	Direct Submission			
JOURNAL	Submitted (09-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>			
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgabs-remail.nih.gov">cgabs-remail.nih.gov</a> Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) BC Cancer Agency, Vancouver, BC, Canada <a href="mailto:info@bcgsc.bc.ca">info@bcgsc.bc.ca</a> Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalios, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.			
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAL Plate: 18 Row: d Column: 24. Location/Qualifiers 1..2627 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:3351791" /tissue type="Eye, retinoblastoma" /clone_lib="NIH MGC 16" /lab_host="DH105-R" /note="Vector: POTB7"			
source				

## ORIGIN

Alignment Scores:				
Pred. No.:	6,11e-86	Length:	2627	
Score:	1003.00	Matches:	233	
Percent Similarity:	53.59%	Conservative:	13	
Best local Similarity:	50.76%	Mismatches:	48	
Query Match:	29.47%	Indels:	166	
DB:	9	Gaps:	10	
US-09-932-678-2 (1-651) x BC006441 (1-2627)				
QY	223	GluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGlu	242	
Db	1509	GAATGTTAGCTTCACTTAAGGATTAGTGATATTTTCCAACTTGGAGCATGAA	1568	
QY	243	IleLeuGluLeuIleGluLysLeuLeuValAsnAlaSerArgGlnGly	262	
Db	1569	ATTCTGGAGCTTATTATTGAAAACTACTCAAGCTGGATGTGAATGTCATCCCGCAGGT	1628	
QY	263	IleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGly	282	
Db	1629	ATTGAGATGCTGAAGAAACAGCAATCAAACTTGTGTGGACAGATTCCACGGAAGA	1688	
QY	283	LeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProGluArg	302	
Db	1689	TGTGTTAATATG-----	1700	
QY	303	LeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu	322	
Db	1700	-----	1700	
QY	323	SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAsp	342	
Db	1700	-----	1700	
QY	343	LeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuProThrHisAlaSerCys	362	
Db	1700	-----	1700	
QY	363	HisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPhe	382	
Db	1701	-----GGATTGCGAGGCAATTT	1718	
QY	383	LeuGluHisLeuTyrLysLysLeuGlnAspProSerAsnProAlaIleArgGlnAla	402	
Db	1719	TTGGAACATCTTTGGAAAACTTCAGGATCCAGTAATCTGCCATCATCAGGAGCT	1778	
QY	403	AlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrVal	422	
Db	1779	GCTGGAAATTAATTGGAAAGCTTTTGGCAAGAGCTAAATTTATTTCTTTATTACTGTA	1838	
QY	423	LysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSer	442	
Db	1839	AAACCATGCTTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCCAGGATTCG	1898	
QY	443	GlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla	462	
Db	1899	GGACAAGGCAATCTCGGATGTTGCTCTCCATGGACATTTTACTCAGCTGCCAGCT	1958	
QY	463	ValPheTyrThrPheValPheArgHisLysGlnLeuSerGlyAsnLeuLysGluGly	482	
Db	1959	GTGTTCTACACCTTTGTTTGTAGACACAAGCAGCTTTTGGCGGAAACCTGTAAGAGAGGT	2018	
QY	483	LeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeu	502	
Db	2019	TTGCAGTATCTCTCAGAGCTCTGAATTTTGGCGGAGTAGTAGTATGAGCCAGCTAAATCCCTG	2078	
QY	503	LysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyr-GlnLe	522	
Db	2079	AAGATTTCCTGCCTCAGTGGTTAACTTTTCTGCAATACACA---AAGATGAAGACT	2135	
QY	522	uValPheCysTyrThrIleLeuGluArgAsnAsnArgGlnMetLeuProValIleArgse	542	

```

Db      2136 TGTG-----GA 2141
QY      542 rThralaglyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPh 562
Db      2142 TATGATGTTGTTGATGGT-----TGCACAACAAT-ATCAATTTATTTATACCAC 2191
QY      562 easPProCysValLeuLys-----568
Db      2192 TGAACGTCGACTTCAAAATGGTTAAGATGGTAAGTTTTATGTTGTATTTTACCATAA 2251
QY      569 -----AtqSerLysLysPheIleAspProIleTyrGlnValTrpGlu-----582
Db      2252 TAAAAAAATTTAGAGGGAAAAACAGTCTGCCTCCACTTTG--ATATGGCAGTGCTA 2308
QY      583 -----AspMetSerAlaGluGluLeuGlnGluPheLysLysProMe 596
Db      2309 ACATCTTCCACCTCCCTCTGCCCTCTGCCCTCCACATCTGGGCAAGCTAAGAAAGCT-- 2366
QY      596 tlyLysAspIleValGluAspGluAspAspPheLeuLysGlyGluValProGlnAs 616
Db      2367 -----GCTGCTCTCTCCTC 2380
QY      616 nAspThrValIleGlyIleThrProSer-----SerPheAspThrHisPheAr 632
Db      2381 TGGCACCACTGGAAATTCATACCAAGCCCTAGCCCTCCACCAGACCCACATTC 2440
QY      632 gSerProSerSerValGlySerProProValLeuTyrMetGlnProSerPro 650
Db      2441 ATCCCCATCCCATCGCATCCCATCCC-----CATCCCCATCCC 2480

RESULT 11
AX713860
LOCUS      AX713860      2493 bp      DNA      linear      PAT 15-APR-2003
DEFINITION      Sequence 544 from Patent EP1293569.
ACCESSION      AX713860
VERSION      AX713860.1 GI:29888786
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1
AUTHORS      Iisogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshikawa,I., Otsuka,M., Nagahari,K. and Masuho,Y.
TITLE      Full-length cDNAs
JOURNAL      Patent: EP 1293569-A 544 19-MAR-2003;
FEATURES      Helix Research Institute (JP) ; Research Association for Biotechnology (JP)
SOURCE      Location/Qualifiers
ORIGIN      1..2493
              /organism="Homo sapiens"
              /mol_type="unassigned DNA"
              /db_xref="taxon:9606"

Alignment Scores:
Pred. No.:      1,1e-85      Length:      2493
Score:          1000.00      Matches:      188
Percent Similarity: 97.45%      Conservative: 3
Best Local Similarity: 95.92%      Mismatches: 5
Query Match:      29.39%      Indels:      0
DB:              6      Gaps:      0

US-09-932-678-2 (1-651) x AX713860 (1-2493)

QY      374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393
Db      703 TATCGACGGATCGCAGAGGAGTATTTGGAAACATCTTTGGAAAAATTCAGGATCCA 762
QY      394 SerAsnProAlaIleIleArGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413

```

```

Db      763 AGTAATCTGCCATCATCAGGAGGCTGCTGAATATATATATGGAAGCTTTTGGCAAGA 822
QY      414 AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeu 433
Db      823 GCTAAATTTATTCCTCTTATTACTGTAAACCATGCTAGATCTTTTGGTTAACTGGCTG 882
QY      434 HisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHis 453
Db      883 CACATATACCTTAATAACAGGATTCGGGAAACAAAGGCAATCTGCAATGTGCTCCCAT 942
QY      454 GlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln 473
Db      943 GGACCATTTTATCTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACACAGCAG 1002
QY      474 LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArg 493
Db      1003 CTTTGGAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTTCAGAGTCTGAATTTGAGCGG 1062
QY      494 IleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePhe 513
Db      1063 ATAGTGATGAGCAGCATAAATCCCTGAAGATTGCTGCTCCCTCAGTGTGTAACTTTT 1122
QY      514 AlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn 533
Db      1123 GCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGTACACCATCATCGAGAGAACAT 1182
QY      534 ArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThr 553
Db      1183 CGCCAGATCTGCCAGTCATTAGGAGTACCACTGGAGGAGACTCAGTCGAGACCTGCACA 1242
QY      554 AsnProLeuAspThrPhePheProPheAspProCysValLeuLysArg 569
Db      1243 AACCCACTGGACACCTTCGTCCTTTGATCCCTGTGCTGCTGAAGAGG 1290

RESULT 12
AX055742
LOCUS      AX055742      2493 bp      mRNA      linear      PRI 01-AUG-2002
DEFINITION      Homo sapiens cDNA FLJ31180 fis, clone KIDNE2000266.
ACCESSION      AX055742
VERSION      AX055742.1 GI:16550546
KEYWORDS      oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1
AUTHORS      Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S., Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S., Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M., Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Iisogai,T.
TITLE      NEDO human cDNA sequencing project
JOURNAL      Unpublished
ORIGIN      2 (bases 1 to 2493)
              Iisogai,T., Otsuki,T. and Sugiyama,T.
              Direct Submission
              Submitted (24-OCT-2001) Takao Iisogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
              NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.
              Location/Qualifiers

```

```

source
1..2493
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="KIDNE200266"
/tissue_type="kidney"
/clone_lib="KIDNE2"
/note="Cloning vector: pME18SFL3"

ORIGIN
Alignment Scores:
Pred. No.: 11e-85 Length: 2493
Score: 1000.00 Matches: 186
Percent Similarity: 97.45% Conservative: 3
Best Local Similarity: 95.94% Mismatches: 5
Query Match: 29.39% Indels: 0
DB: 9 Gaps: 0

US-09-932-678-2 (1-651) x AK055742 (1-2493)
QY 374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLeuGlnAspPro 393
Db : : : : :
703 TATCGCAGGGGATTCGACAGAGGCAATTTTGGAAACATCTTTGGAAAAATTGCAGGATCCA 762
QY 394 SerAsnProAlaIleAlaArgGlnAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413
Db : : : : :
763 AGTAATCTCGCCATCATCAGGAGGCTGCTGGAAATATATTGGAAGCTTTTGGCAAGA 822
QY 414 AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuValAsnTrpLeu 433
Db : : : : :
823 GCTAAATTTATCTCTTATTACTGTAAACCATGCTAGATCTTTTGGTTAACTGGCTG 882
QY 434 HisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHis 453
Db : : : : :
883 CACATATACCTTAATAACAGGATTCGGGAACAAAGGCATCTGCAATGTTGCTCTCCAT 942
QY 454 GlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln 473
Db : : : : :
943 GGACCATTTTACTAGCTGCCAGCTGTGTCTACACCTTTGTTTGTAGACACAAGCAG 1002
QY 474 LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArg 493
Db : : : : :
1003 CTTTGTAGCGGAACCTGAAAGAGTTTGGAGTATCTTCAGAGTCTGAATTTGAGCGG 1062
QY 494 IleValMetSerGlnLeuAsnProLeuLysLysCysLeuProSerValValAsnPhePhe 513
Db : : : : :
1063 ATAGTGATGAGCGACGATAATATCCCTGAAGATTTGCTGCCCTCAGTGGTTAACTTTT 1122
QY 514 AlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn 533
Db : : : : :
1123 GCTGCAATCACAATAAGTACAGCTGCTCTCTCTACACCATCATCGAGAGGAACAAT 1182
QY 534 ArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThr 553
Db : : : : :
1183 CGCCAGATGCTGCCAGTCATTAGGAGTACCACTGGAGGAGACTCAGTGCACACCTGCACA 1242
QY 554 AsnProLeuAspThrPhePheProPheAspProCysValLeuLysArg 569
Db : : : : :
1243 AACCCACTGGACACCTTCGTCCTCTTGTATCCCTGTGTGTGCTGAAGAGG 1290

RESULT 13
AK114871
LOCUS AK114871 2373 bp mRNA linear INV 30-NOV-2002
DEFINITION Ciona intestinalis cDNA, clone:cieg020118, full insert sequence.
ACCESSION AK114871
VERSION AK114871.1 GI:23586599
KEYWORDS FLI CDNA.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
REFERENCE 1
AUTHORS Satou,Y., Yamada,L., Mochizuki,Y., Takatori,N., Kawashima,T.,

```

```

Sasaki,A., Hamaguchi,M., Awazu,S., Yagi,K., Sasakura,Y.,
Nakayama,A., Ishikawa,H., Inaba,K. and Satoh,N.
A cDNA resource from the basal chordate Ciona intestinalis
Genesis 33 (4), 153-154 (2002)
22191024
12203911
2 (bases 1 to 2373)
PUBMED
REFERENCE
AUTHORS Satou,Y. and Satoh,N.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2002) Nori Satoh, Kyoto University, Department of
Zoology; Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
(E-mail:satoh@ascidian.zool.kyoto-u.ac.jp, Tel:81-75-753-4095,
Fax:81-75-705-1113)
COMMENT Ciona intestinalis cDNA Project (URL:
http://ghost.zool.kyoto-u.ac.jp/indexr1.html).
FEATURES
Location/Qualifiers
1..2373
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="cieg020118"

ORIGIN
Alignment Scores:
Pred. No.: 2.78e-85 Length: 2373
Score: 995.50 Matches: 233
Percent Similarity: 56.91% Conservative: 109
Best Local Similarity: 38.77% Mismatches: 211
Query Match: 29.25% Indels: 51
DB: 3 Gaps: 12

US-09-932-678-2 (1-651) x AK114871 (1-2373)
QY 48 LysThrValArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysGlyGlu 67
Db : : : : :
87 AAAATGTGCCGATTTGGAGGAAAATTTCTGATGATTAAGAAAGTTAAAGCAGGTGAT 146
QY 68 ThrAsnAspPheGluLeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGln 87
Db : : : : :
147 ACCAGCACTATCATCTGTCTTCAGCAACTACGGGATCCAAATTAAGTCAAGCCAG 206
QY 88 IleIleAsnTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAspPheGlu 107
Db : : : : :
207 ATAATGGATGGTTTCAAGAGGTTACTTCAATGTCTCATCTTTAGACCAAGATCAAAAT 266
QY 108 GlnLeuIleSerIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGlu 127
Db : : : : :
267 GAATTTGCAATGCTCTGTAAAGTAGTGTGGATGGAAACGAGAAAGTAGTTTCA 326
QY 128 GluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCys 147
Db : : : : :
327 ACATTTGCTGTGTGTGTGTGCTGCTCTGTTCTGCAAAATCTGTTTCATTTAAGAGAGTT 386
QY 148 LeuSerMetIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVal 167
Db : : : : :
387 ATCAAGTCACTGACCCCAAAATCTAATTCAC-----GATTCAAAATCC 428
QY 168 AspValSerAspSerAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHis 187
Db : : : : :
429 AATGCCACAAAGCAACACCCCAAGTAAACAATGGAGGACATTTTGGACAAATTCAC 488
QY 188 ArgAlaLeuGlnIleAlaArgTyrValProSerThrProTrpPheLeuMetProIle 207
Db : : : : :
489 ATTGCTTTTAAAGAAATACTCGGTATCATTTCCAAGGGCATCAGAACCAATGCTACGGTGT 548
QY 208 LeuValGlnLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHis 227
Db : : : : :
549 TTAGTCAAGGC-ATCCCTTTTCATTCGTAAACCAAAATGATTCCTGTGCTGCTACGTCAA 607
QY 228 AsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeu 247
Db : : : : :
608 AACCTCTGTGTCGTCACATCTTACCAGCCTAAACATGAGATCTGCAATACTTGAACCTGCTG 667

```

248 IleGluLysLeuLeuLysLeuAspValAlaSerArgGlnGlyIle-----Glu 264  
::: |||:  
668 ATAAATAATCTCATGATTGGATGTCTCACTCATAGACACGAGATAGAAAATCATGGAA 727  
  
265 AspAlaGluGluThrAlaThrCysGlyGlyThrAspSerThrGluGlyLeuPhe 284  
::: |||:  
728 ARAATTTGAAGAATA-----GACGATCACGAAGAAATTAGTGTGTTTT 769  
  
285 AsnMetAspGluAsp---GluGluThrGluHisgluThrLysAlaGlyProGluArgLeu 303  
|||:  
770 CAATGACGAAGAGATGCAAGAAAAATGAAGAAGAGATAAAA----- 811  
  
304 AspGlnMetValHisProvalAlaGluArgLeuAspIleLeuMetSerLeuValLeuSer 323  
|||:  
812 -----ATGAACAATTCACATTCGCAATAAATTTGAATTCGCAATGCATATATTTGTTTGAA 865  
  
324 TyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeu 343  
|||:  
866 TATTTTACCAGTTTCATACAGGAAGGGGAACACTACAGTTGCAAGAAACCCGAAACATTT 925  
  
344 TyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHis 363  
|||:  
926 TTTCGTGAAATTTGTTTATATATTGACACCGCAATAATGCCAGTACACCAACTTATCCAT 985  
  
364 ValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeu 383  
|||:  
986 GTACAGTTTGGGTGTTTATACCTGCAGCTTAACACAGAGTTTGTGTGAATCGTTTCATC 1045  
  
384 GluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAla 403  
|||:  
1046 GAACATTTGCGCAAGAAAGTTGTAATCTTAACATGCGCAACATTTGACGACAAACCTAC 1105  
  
404 GlyAsnTyrlleGlySerPheLeuAlaArgAlaLysPheIleProLeulleThrValLys 423  
|||:  
1106 ATTGTTTACATGTCAGCTTATTAGCAGTCGCCACCTATATACCATTTATCAACGGTGACA 1165  
  
424 SerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAsp----- 441  
|||:  
1166 GCATGTAITGACTTGATGTTTGGTGTGGGCACACAGATTACTTGAGAGAGGTCACCGAGTTC 1225  
  
442 ---SerGlyThrLysAlaPheCysAspValAlaLeuHisglYrPropheTyrSerAlaCys 460  
|||:  
1226 ATTTTACCAGATGTATATCATTTGTGATATAACACGAGCATGGGACGTTCTACACGATATGC 1285  
  
461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
|||:  
1286 CAACGACATTTCTATGTCTATTATCTTTAAACATCAAGCCATT-----TGGAAATTCAAA 1339  
  
481 GluGlyLeu-GlnTyrLeuGlnSer-LeuAsnPheGluArgIleValMetSerGlnLeuAla 500  
|||:  
1340 ATAGGGTCTATAAATATTTCCACACCTTTAAACTTGGCAACATCTCATAGTTAGTCCACTTA 1399  
  
500 snProLeuLysIleCysLeuProSerValValAsnPheAlaIleThrAsnLys 520  
|||:  
1400 ATCCCTTAAAGCTTGTTTATCATCTGTGGTCAGTTGTTCACCTCTACAAATGAGGCAGC 1459  
  
520 yrGlnLeuValPheCysTyrThrIlelleGluArgAsnAsnArgGlnMetLeuProValI 540  
|||:  
1460 ACCAACTTGTGTA-TGTGACACAATAGT-GAAAGAAACAATTCGACAAATGTTACAGTTG 1517  
  
540 leArgSerThrAlaglyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPheP 560  
|||:  
1518 TT-----GECATGGCACTTTATCTGTAAATAAACCTTCCTTGGTTCATTTT 1562  
  
560 heProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnV 580  
|||:  
1563 TTCCTTTTGCACCTTATCTGTATAAAGGTCTTCTACTTTTTCATCCACCGCATATACAAAG 1622  
  
580 alTrpGluaspMetSerAlaGlu---GluLeuGlnGluPheLysGlyProMetLysLys 599  
|||:  
1623 TGTGGGAGGGGAGTTCGCTCATCTGTAAGATGAAGACATGACGATCCCAATAAAGAAA 1682  
  
599 spIleValGlu-----AspGluAspAspPheLeuLysGlv----- 611

Db	1683	ATATTGAAGATCAGGGGTTATTTCATGAAGATGATACGGCATCAAGGGATCTTTTCGA	1742
Qy	612	--GluValProGlnAsnAspThrValIleGly-----IleThrProSerSerPhe	627
Db	1743	ATCAAGTTCCTCCCTTCCTCATTCACCTCGTTTTCACACATGTTACTCCATCTCTTAC	1801
RESULT 14			
LOCUS	BC055781		
DEFINITION	Mus musculus expressed sequence AL023001, mRNA (cDNA clone IMAGE:6403754), partial cds.		
ACCESSION	BC055781		
VERSION	BC055781.1	GI:33585913	
KEYWORDS			
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; 1 (bases 1 to 2243)		
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyak, S.W., Viallon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
MEDLINE	22388257		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 2243)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-AUG-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library Preparation: M. Bento Soares, University of Iowa cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: <a href="http://www-shgc.stanford.edu">http://www-shgc.stanford.edu</a> Contact: (Dickson, Mark) <a href="mailto:mcd@paxil.stanford.edu">mcd@paxil.stanford.edu</a> Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.		
FEATURES	Location/Qualifiers		
source	1..2243 /organism="Mus musculus" /mol_type="mRNA"		



```

/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6403754"
/tissue_type="Brain, mouse 12.5 dpc"
/clone_lib="NIH BMAP_FDO"
/lab_host="DH10B"
/notes="Vector: pYX-ASC"
<1..2243
/gene="AL023001"
/notes="synonym: Rrn3"
/db_xref="LocusID:106298"
/db_xref="MGI:2146525"
<1..647
/gene="AL023001"
/codon_start=3
/product="AL023001 protein"
/protein_id="AAH55781.1"
/db_xref="GI:33585914"
/db_xref="LocusID:106298"
/db_xref="MGI:2146525"
/translation="RAFCUVALHGFYSACQAVFVTVPRHKQLLSGNLKGLOYLQS
INFERIVQLNPLKICLPQVNVFFAATNKYQLVFCYTIMERNRQMLPVIRTAGG
DSVQCTMPLTFPFDPCVLKRSKKFIDPIYQIWEIDGSAEELQKFKSTKKQVVEDE
DDFLKGVBPQSDTVTGLTPSFDTHFQSPSSVSGPPVLYIPQSPALLTRYD"
27..434
/misc_feature
/gene="AL023001"
/notes="Region: RNA polymerase I specific transcription
initiation factor RRN3. This family consists of several
eukaryotic proteins which are homologous to the yeast RRN3
protein. RRN3 is one of the RRN genes specifically
required for the transcription of rDNA by RNA polymerase I
(pol I) in Saccharomyces cerevisiae"
/db_xref="CDD:pfam05327"

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 6,21e-85 Length: 2243
Score: 991.50 Matches: 187
Percent Similarity: 94.71% Conservative: 10
Best Local Similarity: 89.90% Mismatches: 10
Query Match: 29.14% Indels: 1
DB: 10 Gaps: 1

US-09-932-678-2 (1-651) x BC055781 (1-2243)

QY 445 LysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPhe 464
Db 3 AAGGCTTTTGTGAGCTTGCACTCCATGGACCAATTTATTTCAGCTGCCAAGCTGTGTC 62
QY 465 TyrThrPheValPheArgHisGlyGlnLeuLeuSerGlyAsnLeuLeuGlyGluGlyLeuGln 484
Db 63 TACACTGTTGTTTTAGACACAGCAGCTTTGTAGTGGAACCTTGAAGCAAGGCTACAG 122
QY 485 TyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIle 504
Db 123 TATCTTCAGAGCTCAATTTTGGCGCATTTGTTGAGCCAGCTGACCCCAAGTCAAGATC 182
QY 505 CysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyrGlnLeuValPhe 524
Db 183 TGCCTGCGCGAAGTGGTAAATTTCTTGCTGCTATCACAAATAATACCAGCTGGTGTTC 242
QY 525 CysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAla 544
Db 243 TGCTACACCATCATGGAGAGAAACAGTCCGAGATGCTCCCGTTATCCGAGCAGCAGCT 302
QY 545 GlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspPro 564
Db 303 GGTGGAGACTCCGGGCAGACCTGCACCAACCATCGACACTTCTTCCCTTTGACCCCT 362
QY 565 CysValLeuLysArgSerLysPheIleAspProIleTyrGlnValTyrGluAspMet 584
Db 363 TGTGTGCTTAAGAGGTCAAGAGAGTTCATTGATCTCTTTATCAGATTTGGAGATGGG 422

```

```

QY 585 SerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGlu 604
Db 423 AGTGCTGAAGAGCTTCAGGAGTTTAAAGAAATCTACTAAAAGCAGGATGGAGATGAA 482
QY 605 AspAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrPro 624
Db 483 GATGATGACTTTTGAAGGCGAGGTGCCCCAGAGTGACACAGTGACTGGCCTTACTCCG 542
QY 625 SerSerPheAspThrHisPheArgSerProSerSerSerValGlySerProProValLeu 644
Db 543 AGCTCTTTGATACCACCTTCCAAAGTCTCTCCAGTAGTGGGCTCCCTCTCTGTGCTG 602
QY 645 TyrMetGln---ProSerProLeu 651
Db 603 TATATACCAGGCCAGTCTCCACTC 626

RESULT 15
AK126166
LOCUS
DEFINITION
3645 bp mRNA linear PRI 09-SEP-2003
Homo sapiens cDNA FLJ44178 fis, clone THYMU2037233, highly similar
to Homo sapiens RNA polymerase I transcription factor RRN3 (RRN3).
ACCESSION
AK126166
VERSION
AK126166.1 GI:34532564
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
Arita,M., Mutsashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J.,
Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K.,
Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H.,
Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A.,
Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masubo,Y., Nagai,K.
and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3645)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1..3645
location/Qualifiers
/organism="Homo sapiens"
/db_type="mRNA"
/db_xref="taxon:9606"
/clone="THYMU2037233"
/tissue_type="thymus"
/clone_lib="THYMU2"
/notes="cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:
Pred. No.: 4,16e-84 Length: 3645
Score: 986.00 Matches: 212
Percent Similarity: 62.54% Conservative: 5
Best Local Similarity: 61.10% Mismatches: 17
Query Match: 28.97% Indels: 114
DB: 9 Gaps: 4

US-09-932-678-2 (1-651) x AK126166 (1-3645)

```

Search completed: July 18, 2004, 10:56:28  
Job time : 7164 secs

```
QY      223  GluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGlu 242
Db      2588  GAATGTTACGTTCAATAACTTACTAGGATAGTGATATATTTCCAACTTGAGCATGAA 2647
QY      243  IleLeuGluLeuIleIleGluLysLeuLysLeuAspValAsnAlaSerArgGlnGly 262
Db      2648  ATTCGGAGCTTATTATTGAAAACTACTCAAGCTGGATGTGAATGCATCCCGCAGGCT 2707
QY      263  IleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGly 282
Db      2708  ATTGAAGATGCTGAAGAAACAGCAAACTCAAACTTGTGTGGACACAGATTCCACGGAAGA 2767
QY      283  LeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyProGluArg 302
Db      2768  TTGTTTAATATG----- 2779
QY      303  LeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu 322
Db      2779  ----- 2779
QY      323  SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAsp 342
Db      2779  ----- 2779
QY      343  LeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCys 362
Db      2779  ----- 2779
QY      363  HisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPhe 382
Db      2780  -----GGAATTCGAGAGGCATTT 2797
QY      383  LeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAla 402
Db      2798  TTGGAACATCTTTGGAAAACTTGCAGGATCCAAAGTAATCTGCCATCATCAGCGAGGCT 2857
QY      403  AlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrVal 422
Db      2858  GCTGGAAATTTAATTTGGAAGCTTTTGGCAAGAGCTAAATTTATTTCTTTACTGTA 2917
QY      423  LysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSer 442
Db      2918  AAACCATGCGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATACCGAGGATCG 2977
QY      443  GlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAla 462
Db      2978  GGAACAAAGGCATTCGCGATGTGCTCTCCATGGACATTTTACTCAGCTGCCAAGCT 3037
QY      463  ValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482
Db      3038  GTGTTCTACACCTTTGTTTTTAGCACACAGAGCTTTTGACGCGAAACCTGMAAGAGGT 3097
QY      483  LeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeu 502
Db      3098  TTGCAGTATCCTCAGAGTCTGAATTTTGACGGATAGTAGATGAGCCAGCTAAATCCCCCTG 3157
QY      503  LysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyr-GlnLe 522
Db      3158  AAGATTTGCGCTGCCCTCAGTGGTTAACTTTTGTCTGCAATCACA---AAGATGAAGACT 3214
QY      522  uValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSe 542
Db      3215  TGTG-----GA 3220
QY      542  rThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPh 562
Db      3221  TATGGATGGTGGTGTGATGTT-----TGCAACAACAAAT-ATCAATTTATTTATACCAC 3270
QY      562  eAppProCysValLeuLys 568
Db      3271  TGAACCGTGCACCTTCAAAA 3289
```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 08:21:17 ; Search time 4405 Seconds  
(without alignments)  
4413.230 Million cell updates/sec

Title: US-09-932-678-2  
Perfect score: 3403  
Sequence: 1 MAAPLLHTRLPEDAAASSA.....RSPSSVGVSPVLYMQPSPL 651

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delep 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US0932678/runat\_15072004\_103604\_5218/app\_query.fasta\_1.839  
-DB=EST -OPMT=fastap -SUPFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pcct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0932678.scgn\_1\_3437 @runat\_15072004\_103604\_5218 -NCFU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2966.5	87.2	3457	11	AK080397	AK080397 Mus muscu
2	2742.5	80.6	2136	11	AK080680	AK080680 Mus muscu
3	1755.5	51.6	1800	14	Y16973	Y16973 RNY16973 Ra
4	1290	37.9	896	13	BQ214727	BQ214727 AGENCOURT
5	1242.5	36.5	894	14	CD359064	CD359064 AGENCOURT
6	1223	35.9	889	13	BUS10623	BUS10623 AGENCOURT
7	1208	35.5	828	13	BUE10641	BUE10641 UI-M-FCO-
8	1207.5	35.5	788	12	BG502871	BG502871 60250430
9	1125.5	33.1	996	10	BF203684	BF203684 601866331
10	1120	32.9	754	14	CA511975	CA511975 UI-R-EJ0-
11	1116.5	32.8	769	14	CF738084	CF738084 UI-M-HD0-
12	1111	32.6	682	14	CB150625	CB150625 K-EST0207
13	1105	32.5	733	12	BI561523	BI561523 603256222
14	1100.5	32.3	1037	12	BG428305	BG428305 602498983
15	1094	32.1	757	10	BE734354	BE734354 601565596
16	1048	30.8	651	14	CF908118	CF908118 A0519A10-
17	1040	30.6	671	9	AV703279	AV703279 AV703279
18	1040	30.6	1122	13	BQ066201	BQ066201 AGENCOURT
19	1008.5	29.6	827	13	BUI29543	BUI29543 603118483
20	1003.5	29.5	2659	11	BC037959	BC037959 Homo sapi
21	1000.5	29.4	828	12	BI761160	BI761160 603043657
22	989.5	29.1	736	12	BI464721	BI464721 603202341
23	975	28.7	580	12	BM564978	BM564978 iJ71d11.Y
24	968.5	28.5	878	13	BF743535	BF743535 BX743535
25	968	28.4	609	14	CF902795	CF902795 A0346A07-
26	966	28.4	563	10	AW408066	AW408066 UI-HF-BM0
27	966	28.4	727	12	BI549394	BI549394 603190049
28	964	28.3	551	12	BM786521	BM786521 K-RST0065
29	961.5	28.3	690	9	AJ446028	AJ446028 AJ446028
30	960	28.2	712	13	BY733332	BY733332 BY733332
31	955	28.1	568	9	AL708660	AL708660 DKFZp686K
32	951	27.9	645	13	BQ638571	BQ638571 hd25a12.Y
33	946	27.8	654	14	CF737683	CF737683 UI-M-HD0-
34	941	27.7	571	14	CA563382	CA563382 K0313C08-
35	937.5	27.5	750	9	AJ446022	AJ446022 AJ446022
36	936	27.5	612	13	BQ771414	BQ771414 UI-M-F10-
37	935	27.5	990	12	BG167061	BG167061 602344740
38	928.5	27.3	1045	12	BG297178	BG297178 602395174
39	923.5	27.1	670	13	BY721737	BY721737 BY721737
40	921	27.1	753	14	CB266771	CB266771 1005677 H
41	919.5	27.0	916	13	BX455091	BX455091 BX455091
42	911.5	26.8	703	13	BUT03050	BUT03050 UI-M-PD0-
43	909	26.7	598	14	CF910307	CF910307 A0547H07-
44	902.5	26.5	842	9	AJ446021	AJ446021 AJ446021
45	902	26.5	564	10	AW958173	AW958173 EST370243

# ALIGNMENTS

RESULT 1  
AK080397  
LOCUS  
DEFINITION  
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630092F02 product:RRN3 homolog [Homo sapiens], full insert sequence.  
ACCESSION  
AK080397  
VERSION  
AK080397.1 GI:26099204  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE	1	Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253 MEDLINE PUBMED
REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374 MEDLINE PUBMED
REFERENCE	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 MEDLINE PUBMED
REFERENCE	4	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001) 5
REFERENCE	5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 3457)
REFERENCE	6	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katch, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurikara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, X., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers 1. .3457
FEATURES		source

/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="CS7BL/6J"	
/db_xref="FANTOM_DB:A630092F02"	
/db_xref="MGI:2407489"	
/db_xref="taxon:10090"	
/clone="A630092F02"	
/tissue_type="thymus"	
/clone_lib="RIKEN full-length enriched mouse cDNA library"	
/dev_stage="3 days neonate"	
14. .1987	
/note="PRN3 homolog [Homo sapiens] (SPT Q9NYV6, evidence: FASTY, 87.8%ID, 92%length, match=1791) putative"	
misc_feature	
ORIGIN	
Alignment Scores:	
Pred. No.: 0	Length: 3457
Score: 2966.50	Matches: 570
Percent Similarity: 93.13%	Conservative: 40
Best Local Similarity: 87.02%	Mismatches: 39
Query Match: 87.17%	Indels: 6
DB: 11	Gaps: 2
US-09-932-678-2 (1-651) x AK080397 (1-3457)	
QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAla 20	
Db 14 ATGGGGCGCCCGCTGTTACACGCGCTTGTGGGCGATGTGACGCGACGCGCTTGGC 73	
QY 21 ValLysLysLeuGlyAlaSerArgThrGlyLysSerAsnMetArgAlaLeuGluAsnAsp 40	
Db 74 ACG-----CTGAGTGCCTCGCGACTGGGCTTTTCGATATGCTCGGTAGAGCGAT 127	
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60	
Db 128 TTCTTCAATTTCTCTCGGAAAAAACTGTTGGGTTTGGCGAACTGTGACAGAGCTTTA 187	
QY 61 LeuLysTyrLysGlyGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80	
Db 188 CTGAATATCAAAAGGCGTGAACAAATGCTTAGAGTTGTTGAAGAACACGCTGTCTGAT 247	
QY 81 ProAspLysAspAspGlnIleleAsnTrpLeuLeuGluPheArgSerSerIleMet 100	
Db 248 CTGATATTAAGAGTACACAGATCATTAATCGCTACTAGATTTGTTCTCTGTCATG 307	
QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIlelelelelelelelelelelelele 120	
Db 308 TACTTGACAAAAGACTTTGAGCAACTTATAAACATCATATTGAGATTGCCCTGGTTGAAT 367	
QY 121 ArgSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140	
Db 368 AGAAGTCAGAGAGTGGTGGAGAGTATTTGGCTTTTCTTGGTAATCTTGTCCTGCACAG 427	
QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160	
Db 428 ACTGCTCTCTTAGACCATGTCTCAGCATGATTGGTCTCTCAITTTTCTACTCCCGGATA 487	
QY 161 llelleLysGluGlyAspValAspValSerAspSerAspAspGluAspAspLeuPro 180	
Db 488 ATTGTCAAGAAAGTGGCAGATAGTGTTCAGATTCTGTGATGACGAGATGACATCTTCCT 547	
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIlelelelelelelelelelelelele 200	
Db 548 GCAATTTTGGACACATGTACAGAGCTTGGCAATTAATAACAGATATGTCCTATCGACA 607	
QY 201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArg 220	
Db 608 CCATGGTTTCTAATGCCATACTGGTGGAGAAATTCCTGTTTGGAGGAAGTCCGAGAGA 667	
QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240	
Db 668 ACATTGGAATGTTATGTTTCATTAATTAAGGATAAGTTTATATTATTTCCCAACTTTGAGG 727	

QY	241	HisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAsp-ValAsnAlaSerAr	260
Db	728	CCTGAAATCTCGAGCTGTTATTTGAAAGCTACTCAAGTTAGATGTGAGTGTATCGCG	787
QY	260	gGlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerTh	280
Db	788	GCAGGATATTGAAGATGCTTGAAGAGAAAGCAGCTCAGACTTGTGTGGCAGATACCAC	847
QY	280	rGluGlyLeuPheAsnMetAspGluaspGluGluThrGluHisGluThrLysAlaGlyPr	300
Db	848	AGAAGAGCTGTTTATATATGATGAAGATGAGGACACTGACCTCGAAGAAAGCTGACCA	907
QY	300	oGluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLe	320
Db	908	GGAGCAGCTTAACCAAGATGGCTCATCCACGCAGAGCGCTGAGCGTCTGTGTGCTGT	967
QY	320	uValLeuSerTyrMetLysAspValCysTyrValaspGlyLysValAspAsnGlyLysTh	340
Db	968	GCTGTGTCCTCATVTGAGGATGTCGCGTGTGCACGGTAAAAATGTATAACAATAAAC	1027
QY	340	rLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAl	360
Db	1028	AAAGATTTATACCGTGAICTGATATCCATCTTGACAAACTCTGTGGCCACACATGC	1087
QY	360	aSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGl	380
Db	1088	CTCTCGCCATGTACAGTTCTTTCATGTTTTCCTCTGCAGCTTCAAGTTGGATTTCAGA	1147
QY	380	uAlaPheLeuGluHisLeuTyrLysLysLeuGlnAspProSerAsnProAlaIleLeAr	400
Db	1148	AGCAATTTTGGAAATCTCTGGAAGAAAGTTGCAGATCCAAATAACCCGCCATCATCAG	1207
QY	400	gGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIl	420
Db	1208	GCAGCTGCTCAAATTAATATGGAGCTTTTGGCCAGAGCTAAATTTATTTCTCTTAT	1267
QY	420	eThrValLysSerCysLeuAspLeuLeuValAsnTyrLeuHisIleTyrLeuAsnAngl	440
Db	1268	CACGTGCAAGACATGCCTGGATCTCTTGGTTAACTGCTGCATGATPACCTTTACTAACCA	1327
QY	440	nAspSerGlyThrIlysalPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCy	460
Db	1328	GGATTCCGGAAACAAAGGCTTTTGTGACGTGACATCCATGGACCAATTTATTTCAGCTG	1387
QY	460	sGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLy	480
Db	1388	CCAAGCTGCTCTACACTGTGTTTATGACACAGAGCAGCTTTTGAAGTGAACCTGAAR	1447
QY	480	sGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAs	500
Db	1448	GCAAGGTCTACAGTATCTTCAGAGTCTAAATTTTGAAGCGATTTGTGTGAGCGACCTGAA	1507
QY	500	nProLeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTy	520
Db	1508	CCACCTGAAGATCTGCTCGCCAGTGGTTAATTTCTTTGTGCTATCACAAATAAATA	1567
QY	520	rGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIl	540
Db	1568	CCAGCTGGTGTCTGTTACCACTATGAGAGAGAAACATGTCGCGACATGCTCCCGCTTAT	1627
QY	540	eArgSerThrAlaGlyClyAspSerValGlnIleCysThrAsnProLeuAspThrPhePh	560
Db	1628	CCGACAGACCGCTGGTGGAGACTCCGTGCAGACCTGCACCAACCCACTGGACACTTCTT	1687
QY	560	eProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVa	580
Db	1688	CCCTTTTGACCTTGTGCTTAAGAGGTCAAAGAAGTTCAATTGATCTTATTATTCAGAT	1747
QY	580	lTyrGluAsp-MetSerAlaGluGluLeuGln-GluPheLysLysProMetLysLysAsp	599
Db	1748	TTGGGAAGATGGGAGTGTGAAGAGCTTCAGGAGTTTAAAGAAATCTACTATAAAGAG	1807

QY	600	lleValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspThrVal	619
Db	1808	CTAGTCGGAGGATGAAGATGATGACTTTTGAAGGCGAGGTGCCCCAGATGACACAGTG	1867
QY	620	lleGlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerValGly	639
Db	1868	ACTGCGCCCTACTCCGAGCTCTTTGATACCCACTTCCAAAGTCTCCAGTAGTGTGGC	1927
QY	640	SerProValLeuTyrMetGln--ProSerProLeu	651
Db	1928	TCCCTCTCTGCTGTATATACACAGCGCAGTCTCCACTC	1966
RESULT 2			
LOCUS	AK080680	2136 bp mRNA linear	HTC 19-SEP-2003
DEFINITION	Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched library, clone:A830089A07 product:RRN3 homolog [Homo sapiens], full insert sequence.		
ACCESSION	AK080680.1	GI:26099414	
VERSION	AK080680		
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE			
AUTHORS	Carninci,P. and Hayashizaki,Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	92279253		
PUBMED	10349636		
REFERENCE			
AUTHORS	Carninci,P. and Hayashizaki,Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE			
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE			
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
MEDLINE	11076861		
PUBMED			
REFERENCE			
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
JOURNAL	Nature 420, 563-573 (2002)		
MEDLINE	6 (bases 1 to 2136)		
PUBMED			
REFERENCE			
AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohnato,N., Okazaki,Y., Saijo,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,		

Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.  
 Direct Submission  
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: <http://genome.gsc.riken.go.jp/>  
 URL: <http://fantom.gsc.riken.go.jp/>  
 Location/Qualifiers  
 1. 2136  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /st-ain="C57BL/6J"  
 /db\_xref="PANTOM\_DB:A830089A07"  
 /db\_xref="WGI:2409380"  
 /db\_xref="taxon:10090"  
 /clone="A830089A07"  
 /tissue\_type="cortex"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="10 days neonate"  
 40. 1905  
 /note="RRN3 homolog [Homo sapiens] (SPTR|Q9NVV6, evidence: FASTV, 87.8%ID, 92%length, match=1791)  
 putative"  
 2110. 2115  
 /note="putative"  
 2136  
 /note="putative"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 2136  
 Score: 2742.50 Matches: 549  
 Percent Similarity: 86.51% Conservative: 41  
 Best Local Similarity: 80.50% Mismatches: 56  
 Query Match: 80.59% Indels: 38  
 DE: 11 Gaps: 7  
 US-09-932-678-2 (1-651) x AK080680 (1-2136)  
 QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaSerSerAla 20  
 Db 40 ATGGCGGCCCGCTGCTTACACGCGCTTGTCCGCGATGTGACGCGCGCGCTCTGCC 99  
 QY 21 ValLysLeuGlyAlaSerArgThrGlyLeuSerAsnMetArgAlaLeuGluAsnAsp 40  
 Db 100 ACG-----CTGAGTCGCTGCGGACTGGGCTTCCGATATGCTCGGTTAGAGCGCAT 153  
 QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyThrValThrGluValLeu 60  
 Db 154 TTCTTCAATCTCTCCGAAAAAACTGTTCGTTGGCGGAACCTGTGACAGAAGCTT 213  
 QY 61 LeuLysTyrlsLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80  
 Db 214 CTGAATACAAAAGGCTGAAACAAATGACTTAGAGTTGTTGAAGAACCAAGCTGCTCAT 273  
 QY 81 ProAspIleLysAspAspGlnIleLeuAsnTrpLeuGluGluPheArgSerIleMet 100  
 Db 274 CCTGATATAAGGATGACCAAGATCAATTAAGTCTAGAAATTCGTTCTCTGTCATG 333  
 QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleLeuArgLeuProTrpLeuAsn 120

Db 334 TACTTGACAAAAGACTTTTGAGCAACTTATATAACATCATATTGAGATTGCCCTGGTGAAT 393  
 QY 121 ArgSerGlnThrValValGluGluGlyLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140  
 Db 394 AGAAGTCAGAGAGTGGGGAGGATATTGGCTTTTCTTGGTAATCTTGTGTGCACAG 453  
 QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160  
 Db 454 ACTGTCTTCTTAGACCATGCTCAGCATGATTGGCTCTCATTTTGTACCTCCCGAGTA 513  
 QY 161 IleIleLysGluGlyAspValAspValSerAspSerAspGluAspAsnLeuPro 180  
 Db 514 ATTGTCAAGGAAGTGGCATAGATGTTTCAGATTCTGATGACGAAGTGAACAATCTTCT 573  
 QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleLeuAlaArgTyrValProSerThr 200  
 Db 574 GCAATTTTGTACATGTCACAGAGCCTTGGAAATATATAACAGATATGTCCATCGACA 633  
 QY 201 ProTrpPheLeuMetProIleLeuValGlyLysPhePropheValArgLysSerGluArg 220  
 Db 634 CCATGGTTTCTAATGCCAATACTGTCGAGAAATCCCGTTTGTGAGGAAGTCCGAGAGA 693  
 QY 221 ThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuArg 240  
 Db 694 ACATTGGAATGTTATGTTTCACTATTAAGGATAAGTTTATATATTTCCCAACTTTGAGG 753  
 QY 241 HisGluIleLeuGluLeuIleLeuLysLeuLeuLysLeuAspValAsnAlaSerArg 260  
 Db 754 CGTGAATTCGGAGCTTGTATTGAAAGCTACTCAAGTTAGATGTGAGTGTATCGCG 813  
 QY 261 GlnGlyIleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280  
 Db 814 CAGGATATTGAAGTGTCTGAAGAGAAAGCAGCTCAGACTTGTGTGGGACAGATACCACA 873  
 QY 281 GluGlyLeuPheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyPro 300  
 Db 874 GAAGGACTGTTTAAATATGATGATGAAGATGAGGACACTGACCCCTGGAAGAAAGCTGACCAG 933  
 QY 301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeu 320  
 Db 934 GAGCAGCCTAACCAATGGCTCATCCACCGCAGAGCGCTGAGCGTCTCTGTGTGCTTG 993  
 QY 321 ValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThr 340  
 Db 994 CTGCTGTCTCTACATTGAGGATGCTCTGCCGTGTGCACGGTAAATAATGATAACATAAACA 1053  
 QY 341 LysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAla 360  
 Db 1054 AAGGATTTATACCGTATCTGATATCCATCTTTGACAAACTTCTGTGCCACACATGCC 1113  
 QY 361 SerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380  
 Db 1114 TCTGCCATGTACAGTCTTCTCATGTTTTCCTCTGCGCTTCAAGTTGGATTTGCAGAA 1173  
 QY 381 AlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleLeuArg 400  
 Db 1174 GCATTTTGGAAACATCTCTGGAATAAGTTGCGAGATCCAAATAACCCCGCATCATCAGG 1233  
 QY 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle 420  
 Db 1234 CAGGCTGCTGCAAAATATATATTGGAGCTTTTGGCCAGAGCTAAATTTATCTCTTATC 1293  
 QY 421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGln 440  
 Db 1294 ACTGTGAACACATGCTGGATCTCTTGGTTAACTGGCTGCACATGTACCTTACTAACCCAG 1353  
 QY 441 AspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCys 460  
 Db 1354 GATTTCGGGAACAAGGCTTTTGTGACGTTGCACT-CATGGACCATTTTATTAGCCCTGC 1412  
 QY 461 GlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLys 480  
 Db 1413 CAAGCTGTGTTCTACACTGTTGTTTTTAGACACAGCAGCTTTTGAGTGGAACTTGAAG 1472

```

QY 481 GluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500
Db 1473 CAAGGCTACAGTATCTTCAGAGTCTAAATTTTGGCGCATTTGTGTGAGCCAGCTGAAC 1532
QY 501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyr 520
Db 1533 CCACCTGAAGATCTGCCCTGCGGAGTGGTAAATTTCTTGCTGCTATCAATAAATAC 1592
QY 521 GlnLeuValPheCysTyrThrIleLeuGluArgAsnAsnArgGlnMetLeuProValIle 540
Db 1593 CAGCTGGTCTTCTGCTACACCATCATGAGAGGAAACAGTCCGAGATGCTCCCGTATC 1652
QY 541 ArgSer-ThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePh 560
Db 1653 CCAGCCACCGCTGGTGGAGACTCCGTGAGACCTGCACCAACCCACTGGACACTTTCTT 1712
QY 560 eProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnVa 580
Db 1713 CCCCTTTGACCTTGTGTCTTAAGAGGTCAAGAGATTCAATGATCCTATTATCAGAT 1772
QY 580 ITrpGluAspMetSerAlaGluGlnGluPheLysLysProMetLysLys----- 598
Db 1773 TTGGGAAGATGGAGTGTCTGAAGAGCTTCAGGAGTTTAAAGAAATCTACTAAAAAGGAATC 1832
QY 599 -----AspIleValGluAspGluAsp-- 605
Db 1833 TCCTGCTGGAAGAGCAATGSCACCATCTCTGAAGGGAACTGTCTCCAGAGAGATCT 1892
QY 606 -AspAspPheLeuLysGlyGluVal---ProGlnAsnAspThrValIle----- 620
Db 1893 CAATGAGAC-CTAAAAAACTTGTGAGACCAAGATGAGACTAAAGCCTAAATATGAT 1951
QY 621 -----GlyIleThrProSer-SerPheAspThrHisP 631
Db 1952 GGGTAGGATCCAAACAGCAGAAATAAGGGCTCAACCGGATTTATCTGGCAACCGGAT 2011
QY 631 heArgSerProSerSerValGlySerProProValLeuTyrMetGlnProSerProL 651
Db 2012 TTCTCTCCCTCTCTCTCTCTC---TCTCCC-----CTCTCCCTCTCTCCCTCTCGCC 2062
QY 651 eu 651
Db 2063 TC 2064

RESULT 3
Y16973
LOCUS
DEFINITION
RNV16973 Rat liver ESTs (E.Olivier) Rattus norvegicus cDNA clone
ACCESSION
Y16973, mRNA sequence.
VERSION
Y16973.2 GI:4727035
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1. (bases 1 to 1800)
Olivier,E., Soury,E., Risler,J.L., Smith,F., Schneider,K.,
Lochner,K., Jouzeau,J.Y., Fey,G. and Salier,J.P.
A novel set of hepatic mRNAs preferentially expressed during an
acute inflammation in rat represents mostly intracellular proteins
Genomics 57 (3), 352-364 (1999)
99263497
MEDLINE
10329001
PUBMED
COMMENT
On Apr 2, 1998 this sequence version replaced gi:3006069.
U78 INSDM
543 chemin de la Breteque, 76233 Bois-Guillaume, France
POLYA-No.
Location/Qualifiers
1. .1800
FEATURES
source

```

```

/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="IRL573"
/tissue_type="liver"
/dev_stage="adult"
/clone_lib="Rat liver ESTs (E.Olivier)"
/note="Organ: liver; Vector: lambda ZAPII; Library
construction: Ripberger, J.A. et al. J. Biol. Chem. (1995)
270(50):29998-30006"

ORIGIN

Alignment Scores:
Pred. No.: 6,44e-200 Length: 1800
Score: 1755.50 Matches: 451
Percent Similarity: 80.9% Conservative: 51
Best Local Similarity: 72.74% Mismatches: 98
Query Match: 51.59% Indels: 42
DB: 14 Gaps: 15

US-09-932-678-2 (1-651) x Y16973 (1-1800)
QY 24 LeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPheAsn 43
Db 26 CTGAGTGGTGGCGACTGGGCTTTCAAAACATGCTTGCATTAGAGAGTGATTTC----- 79
QY 44 SerProArgLysThrValArgPheGlyGlyThrValThrGluValLeuLysTyr 63
Db 80 TCAGTCTCCAGAAACATGTCGGITGG---TGGACTGTGACAGACGCTCTTACTGAGT--- 133
QY 64 LysLysGlyGluThrAsnAspPheGluLeuLysAsnGlnLeuLeuAspProAspIle 83
Db 134 ACAAAGGGGTGAACAAATGACTTAGAGTTGTTGAAGACAG-CYCTCTGATCTCATATA 192
QY 84 LysAspAspGlnIleIleAsnTrpLeuGluPheArgSerSerIleMetTyrLeuThr 103
Db 193 GAGGATGATCAGATCATTAACCTGGCTGCTGGAATTCGTCTCTGTCATGACTGACATA 252
QY 104 LysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGln 123
Db 253 AAAGACTTTGAGCACTTATTAACATCATATTGAGATTGCCCTGGTTGAATAGAGTACAG 312
QY 124 ThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPhe 143
Db 313 ACGGTGGTGGAGGAGTATTTGGCTTTTCTTGGTAACTTTGTGTGACACAGACTGCTCTC 372
QY 144 LeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgValIleIleLys 163
Db 373 CTTAGACCATGCTCTCAGCATGATTGCTTCTCATTTTGTACCTCCCGAGTTATTGTCAG 432
QY 164 GluGlyAspValSerAspSerAspGluAspAspAsnLeuProAlaAsnPhe 183
Db 433 GAAGGTGCATGATGT-TCAGAGTCTGAGATGAAGATGAACATCTTCCTGCAAAATTT 491
QY 184 AspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPhe 203
Db 492 GACACATGTCACAAAGCCTTCGAAATATAACAAGATATGTCCTTCGACACCATGGTTT 551
QY 204 LeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGlu 223
Db 552 CTAATGCCAATACCTGGTAGAAAAATTTCCATTTGTAGGAAATTCAGAGAACACACAG 611
QY 224 CysTyrValHisAsnLeuArgIleSerValTyrPheProThrLeuArgHisGluIle 243
Db 612 TGTATGTTTATAACCTATTAAAGATAGTTTATTTCACACGTTGAGCGGTGAATTT 671
QY 244 LeuGluLeuIleIleGluLysLeuLysLeuAspValAlaSerArgGlnGlyIle 263
Db 672 CTCGAGCTTATTATTGAAGTCA-CTCAAGTTAGATGTGAGTGTGCGGACAGATATT 730
QY 264 GluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeu 283
Db ::::|||||

```

Db 731 GACGATGCTGAAGAGACAGCAGCTCAGACTTATGTGGGACAGACGCCGAGAGGACTG 790  
QY 284 PheAsnMetAspGluAspGluThrGluHisGluThrLysAlaGlyProGluArgLeu 303  
Db 791 TT-AATATGGATGA-GATGAGGA-ACTGACCCCTGAGAGAAAGCTCAGCAGAGCGCCTG 847  
QY 304 AspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSer 323  
Db 848 GC-CGATGGCTCATCCACTGCAGAGCGCTGGACATCTCTGCTGCTGCTCTCC 906  
QY 324 TyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeu 343  
Db 907 TACATTGACGATGCTGCGCATGTGC--AGTAAATTTGATAACAATAAAACAAGA--TTA 962  
QY 344 TyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHis 363  
Db 963 TATCGTGATCTGATATCTATCTTG---ACAACGTGTGTGCCACACATGCTC-TGCCAT 1018  
QY 364 ValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeu 383  
Db 1019 GTACAGTTCTTCAGTTTTCCTTTCAGCT---CAAAATGGGGTTTCGACAGATTTTG 1075  
QY 384 GluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleLeuArgGlnAlaAla 403  
Db 1076 GAACATCTC---TGAAAAGTGCAGGATCA-AATAAGCCTGTCTCATCATCAGCAGCTGCTG 1131  
QY 404 GlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLys 423  
Db 1132 ---AATATACATG---AGCTTTTGGC-AGAACTTAAATTAAT-CCTCTTATCATCTGGAAG 1183  
QY 424 SerCysLeuAspLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGly 443  
Db 1184 TCATGCTGGATCTTTGGGT-ACTGGCTGCATG-TACCTTAATAGCCAG-----GTT 1235  
QY 444 ThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerLysGlnAlaVal 463  
Db 1236 CGGACAAAGGCTTGTGATGTGC---ACTCATGGACCATTTATCAGC---TGCCAAAGCTGTG 1290  
QY 464 PheTyrThrPheValPheArgHisLysGlnLeuLeuSerClyAsnLeuLysGluGlyLeu 483  
Db 1291 TC-TACACT---GTGTTTAGACACAGCAGCTTTGAGTGCACCT-----GAGGAAGT-CTA 1339  
QY 484 GlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLys 503  
Db 1340 CGGTATCTCAGAGTCTAATTTTGAACGATTTGTCTGAGCCAGCTGAACCCACTGAAG 1399  
QY 504 IleCysLeuProSerValValAsnPheAlaAlaIleThrAsnLysTyrGlnLeuVal 523  
Db 1400 ATCTGCTGCCAGCTGTGTATCTCTGCTGCATACAATA-----ATACAGCTGTGT 1450  
QY 524 PheCysTyrThrIleLeuGluArgAsnAsnArgGlnMetLeuProValIleArgSerThr 543  
Db 1451 CTGCTAC---ATCATGGAGAGAC-AATCGCCAGATGCTCCTGTATTATCG--AGTACA 1504  
QY 544 AlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPheAsp 563  
Db 1505 GCTGTGTGAGACTCTGTGCAGACTGTCACCAACCCATTTGACACTTTCTTCCCTTGAC 1564  
QY 564 ProCysValLeuLysArgSerLysPheIleAspProIleTyrGlnValTrpGluAsp 583  
Db 1565 CCTGTGTATGTTAAGAGTCAAGAGTTTCATGATCTCTCTATCAGGCTTGGGAGAC 1624  
QY 584 MetSerAlaGluGlnLeuGlnPheLysLysProMetLysLysAspIleValGluAsp 603  
Db 1625 ATGAGTGTAGAGCTGAGGAGCTTAGAACCCACT-AGAAAGGAGCTAGTGGAGAT 1683  
QY 604 GluAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThr 623  
Db 1684 GAAGACGATGACATTTTGAAGGCGAGTGCC---CAGAGTCACACAGTGGCTGGCCTTACC 1740  
QY 624 ProSerSerPheAspThrHisPheArgSerProSerSerValGlySerProVal 643  
Db 1741 CCGAACTCTTTGACACCCACTTTTCGAGT-CCTTCCAGTAGTGTGGCTCTCCCTCTCTGTG 1799

## RESULT 4

BQ214727

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

## FEATURES

source

1..896

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6062849"

/tissue\_type="melanotic melanoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH MGC 72"

/notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."

## ORIGIN

Alignment Scores:

Prod. No.: 3 41e-144

Score: 1290.00

Percent Similarity: 91.22%

Best Local Similarity: 89.53%

Query Match: 37.91%

DB: 13

US-09-932-678-2 (1-651) x BQ214727 (1-896)

QY 314 LeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGly 333  
Db 23 GTGGACATCTGATGCTTTGGTTTGTCTCATGAAGGATGCTGCTATGATAGTGT 82  
QY 334 LysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLys 353  
Db 83 AAGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAA 142  
QY 354 LeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSer 373  
Db 143 CTCTGTGTGCCACCCATGCTCTCCATGATGATACAGTTTTTCATGTTTACCTCTGTAGT 202  
QY 374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393  
Db 203 TTCAAAATGGGATTCGACAGGCAATTTTGGAAACATCTCTGGAACAAATTCAGGACCCA 262  
QY 394 SerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413  
Db 263 AGTAATCTCTGCCATCATCAGGCGGCTGCTGGAAATTAATTATTTGGAAGCTTTTGGCAGA 322



```

QY 414 AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeu 433
Db GCTAANTTTATTCCTCTTATTACTGTAAATCATGCCGTAGATCTTTGGTTAACTGGCTG 382
QY 434 HisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHis 453
Db CACATATACCTTAATAACAGGATTCGGGAACAAGGCATCTCGGATGTTGCTCTCCAT 442
QY 454 GlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln 473
Db GGACCAATTTTACTCAGCCCTGCCAGCTGTGTCTACACCTTTGTTTTAGACACAAGCAG 502
QY 474 LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArg 493
Db CTTTGTGACGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGG 562
QY 494 IleValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePhe 513
Db ATAGTGATGAGCCAGCTAAATCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTT 622
QY 514 AlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAs 533
Db GCTGCAATCACAAATAATACACCTGCTCTCTGTACACCCATCAATTGAGAGGAACA 682
QY 533 nArgGlnMetLeuProValIleArgSerThrAla-GlyGlyAspSerValGlnIleCysT 553
Db TCGCCAAATGCTGCCCGTCATTAAGATGACCGCTGGGAGGAGACTCAGTGCATAATCTGCA 742
QY 553 hrAsnProLeuAspThr-PhePheProPhe-AspProCysValLeuLysArgSer---LY 571
Db AAAACCCGCTGCACACCTTTCTTCCCTTTGATCCCTGTGTGCTGGAAGAGGTCCAAA 802
QY 571 sLysPheIle-AspProIle---TyrGlnValTrpGluAspMetSerAla----- 586
Db AAAATTCATGTCATCTATTAATCCGGGGGGGGGAAAAAATGAGTGCCTGAAAAAACCT 862
QY 587 --GluGluLeuGlnGluPheLysLysProMetLysLys 598
Db AACGGGAATTCAGAA-----AACCCATTCAGAAAAA 894

RESULT 5
CD359064
LOCUS AGENECOURT 14303450 NIH_MGC_180 Homo sapiens cDNA clone
DEFINITION IMAGE:30389999 5', mRNA sequence.
ACCESSION CD359064
VERSION CD359064.1 GI:31130475
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 894)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM464 row: f column: 24
High quality sequence stop: 506.
Location/Qualifiers
1..894
/organism="Homo sapiens"

```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30389999"
/lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
/clone_lib="NIH_MGC_180"
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
Site 2: EcoRV (destroyed); Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH_MGC
Library."

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 1,84e-138 Length: 894
Score: 1242.50 Matches: 256
Percent Similarity: 86.89% Conservative: 9
Best Local Similarity: 83.93% Mismatches: 20
Query Match: 36.51% Indels: 20
DB: 14 Gaps: 6

US-09-932-678-2 (1-651) x CD359064 (1-894)

QY 362 CysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAla 381
Db 1 TGCATGTACAGATTTTTCATGTTTACCTCTGTAGTTTCAAAATGGGATTCGCAGAGCA 60
QY 382 PheLeuGluHisLeuTrpLysLeuGlnAspProSerAsnProAlaIleIleArgGln 401
Db 61 TTTTGGAACTCTCTGGAAAAATTCGAGGACCCAGTAATCTCGCCATCATCAGGCGAG 120
QY 402 AlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThr 421
Db 121 GCTGCTCGAAATTAATATGGAAGCTTTTTCGCAAGAGCTAAATTTATTCCTTTACT 180
QY 422 VallySerCysLeuAspLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAsp 441
Db 181 GTAAAAATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAAATACAGGAT 240
QY 442 SerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGln 461
Db 241 TCGGGAACAAGGCAATCTCGATGTTGCTCTCCATGACCATTTTACTCGAGCCTGCCAA 300
QY 462 AlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGlu 481
Db 301 GCTGTGTTCTACACCTTTGTTTAGACACAGCAGCTTTTGACGGGAAACCTGAAAGAA 360
QY 482 GlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPro 501
Db 361 GGTTCAGATATCTTCAGAGTCTGAATTTTGAGCGGATAGTAGAGCAGCTAAATCCC 420
QY 502 LeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsnLysTyrGln 521
Db 421 CTGAGATTTGCTGCGCCCTCAGTGGTTAACTTTTCTGCAATCACAATAAGTACCGAG 480
QY 522 LeuValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArg 541
Db 481 CTGCTCTTCTGCTACACCATCTGTGAGAGGAACAATGCCAGATGCTGCCAGCATTAGG 540
QY 542 SerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhePro 561
Db 541 AGTACCCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTCCC 600
QY 562 PheAsp-ProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGlnValTr 581
Db 601 TTTGATCCCTGTGTCTGGAGAGGACAAAGAAATTCATTGATCCTATTATTCAGGTGTG 660
QY 581 pGluAspMetSerAlaGluGlnLeuGlnPhe-LysLysProMet-LysLysAspIle 600
Db 661 GGAAGACATGAATGCTGAAAGCTACAGGAATTCAGAAACCCATGAAAGAGGACCTA 720
QY 601 Val-----GluAspGluAspAspPheLeuLysGly---GluValProGlnAsn--- 616

```

Db 721 GTGGGAAGATGACGAATGATGAACTTTCTCTGAAAGGCGGAAAGTGCCTCCCAAGAAAGG 780  
 Qy 617 -----AspThrValIleGlyThrProSerSerPheAspThr--HisPhe---A 532  
 Db 781 AAACCCGTGATGTGGAATTCACACCCAAACCTCCCTTTTTCACACCGCCATTTTCCCA 840  
 Qy 632 tGSerProSerSerSer-----ValGlySerProProValLeuTyMetGlnProSerP 650  
 Db 841 GAAACCCCTTCTTAAATAAGTGGTGGGGCGCCCTCCC-----CCACCC 885  
 Qy 650 toLeu 651  
 Db 886 CGGTG 890

RESULT 6  
 BU510623  
 LOCUS AGENCOURT\_10120114 NIH\_MGC\_134 Mus musculus cDNA clone EST 12-SEP-2002  
 DEFINITION IMAGE:6505326 5', mRNA sequence.

ACCESSION BU510623  
 VERSION BU510623.1 GI:22816856  
 KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 889)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)

Tissue Procurement: Dr. David Rowe

cDNA library Preparation: Invitrogen Corp

cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM1466 row: a column: 07

High quality sequence start: 24

High quality sequence stop: 646.

Location/Qualifiers

## FEATURES

source

1..889  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6505326"  
 /tissue\_type="undifferentiated limb"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_134"  
 /note="Vector: PCMV-SPORT6.1; Site.1: EcoRV; Site.2: NotI;  
 Cloned unidirectionally. Primer: Oligo dt. Average insert  
 size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:  
 this is a NIH\_MGC Library."

## ORIGIN

Alignment Scores:  
 Pred. No.: 4,12e-136 Length: 889  
 Score: 1223.00 Matches: 244  
 Percent Similarity: 88.93% Conservative: 13  
 Best Local Similarity: 84.43% Mismatches: 28  
 Query Match: 35.94% Indels: 4  
 DB: 13 Gaps: 1

US-09-932-678-2 (1-651) x BU510623 (1-889)

Qy 4 ProLeuLeuHisThrArgLeuProGlyAspAlaalaSerSerAlaValLysLys 23  
 Db 15 CCAGCTGCTCCACTTCTGTGGGGGATGTGACGGCGGCGCTCGCCAG-----68  
 Qy 24 LeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPheAsn 43

Db 69 CTGAGTGCCTCGCGGACTGGGCTTTCCGATATGCTCGGCTTAGAGAGCGATTCTTCAAT 128  
 Qy 44 SerProArgGlyThrValArgPheGlyGlyThrValThrGluValLeuLeuLysTyr 63  
 Db 129 TCTCTCTCGAATAAACTGTTGCGTTTGGCGAAGTGTGACAGAAGTCTTACTGAAATAC 188  
 Qy 64 LysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIle 83  
 Db 199 AAAAGGGGTGAACAAATGACTTAGAGTTGTTGAAGAACCGAGCTGTCTGATCCTGATA 248  
 Qy 84 LysAspAspGlnIleLeuAsnTrpLeuGluPheArgSerSerIleMetTyrLeuThr 103  
 Db 249 AAGGATGACCAATCAATTAAGTGGCTACTAGAAATTCGTTCTCTCATGTACTGTACA 308  
 Qy 104 LysAspPheGluGlnLeuLeuSerIleLeuArgLeuProTrpLeuAsnArgSerGln 123  
 Db 309 AAAGACTTTGAGCAACTTATATAACATCAATTCAGATTGGCCCTGGTTGATAGAGTCA 368  
 Qy 124 ThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPhe 143  
 Db 369 AGAGTGGTGGAGGAGTATTGGCTTTCTTGTAATCTTGTCTGTCACAGACTGTCTTC 428  
 Qy 144 LeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgValIleLys 163  
 Db 429 CTTAGACCATGCTCAGCATGATTCGCTCTCAATTTGTACCTCCCGAGTAATGTCAAG 488  
 Qy 164 GluGlyAspValAspValSerAspAspGluAspAspAsnLeuProAlaAsnPhe 183  
 Db 489 GAAGTGGCATAGATGTTTCAGATTCTGATGACGAGATGACAACTCTCTCTGCAATTTT 548  
 Qy 184 AspThrCysHisArgAlaLeuGlnIleAlaArgTyrValProSerThrProTrpPhe 203  
 Db 549 GACACATGTCACAGAGCCTTGCAATAATAACAAGATATGTCCTCCATCGACACCATGTT 608  
 Qy 204 LeuMetProIleLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGlu 223  
 Db 609 CTAATGCCCAATACTGGTGAGAAATTCGCGTTTGGAGAAATTCGAGAGAACATTGGAA 668  
 Qy 224 CysTyrValHisAsnLeuArgIleSerValTyrPheProThrLeuArgHisGluIle 243  
 Db 669 TGTTATGTCATACTTATTAGGATAGTTTATTTATTTCCCACTTGAGCGTGAAT 728  
 Qy 244 LeuGluLeuIleIleGluLysLeuLeuAspValAsnAlaSerArgGlnGlyIle 263  
 Db 729 CTGGAGCTTGNTATTGAAAGCTACTCAAGTTAGATGTGAGTGTATCGCGGAGGATATT 788  
 Qy 264 GluAspAlaGlu-GluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGly 283  
 Db 789 GAAGATGCTGAAAGAGAAAGCAGCTCAAACTTGTGTGGGACAGATACCCCGAAAGACT 848  
 Qy 283 u-PheAsnMetAspGluAspGlu 290  
 Db 849 GGTATTATGATGATGAGATGAA 871

## RESULT 7

BU610641

LOCUS

DEFINITION

UI-M-FCO-caq-n-01-0-UI.r1 NIH\_BMAP\_FCO Mus musculus cDNA clone

UI-M-FCO-caq-n-01-0-UI 5', mRNA sequence.

ACCESSION

BU610641

VERSION

BU610641.1 GI:23276856

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 828)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)

Tissue Procurement: Dr. James Lin, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5,

Location/Qualifiers

1. .828

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CS7BL/6"

/db\_xref="taxon:10090"

/clone="UI-M-FC0-caq-n-01-0-UI"

/tissue\_type="whole brain"

/dev\_stage="embryo 12.5 dpc"

/lab\_host="DH10B (T1 phage resistant)"

/clone\_lib="NIH\_BMAP\_FC0"

/note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dr primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with EcoR I adaptor, digested with Not I, and then  
cloned directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA  
tail, is TCAGAGAGCC. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP). 'Gene Discovery in the Developing Mouse Nervous  
System', supported by National Institutes of Mental Health  
(NIMH), Hemin Chin, Ph.D., program coordinator."

# ORIGIN

## Alignment Scores:

Pred. No.: 2,398-134 Length: 828  
Score: 1208.00 Matches: 233  
Percent Similarity: 91.27% Conservative: 18  
Best Local Similarity: 84.73% Mismatches: 23  
Query Match: 35.50% Indels: 2  
DB: 13 Gaps: 0

US-09-932-678-2 (1-651) x BU610641 (1-828)

QY 179 LeuProAlaAsnPhaAspThrCysHisArgAlaLeuGlnIleAlaArgTyrValPro 198  
DB 1 CTTCCTGCATTTTTCACACATGTACAGAGCTTGCATTAATATACAGATATGTCCCA 60  
QY 199 SerThrProTTPheLeuMetProIleLeu-ValGluLysPheProPheValArgLysE 218  
DB 61 TCGACACCATGTTTCTAATGCAATCTACTGGGTGAGAAATCCCGTTTGTGAGGAAGTC 120  
QY 218 cLluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProTh 238  
DB 121 CGAGAGACATTGGAAATGTATGTTCATACATTTATTAAGGATAAGTTTATATTTCCCAAC 180  
QY 238 rLeuArgHisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuAspValAsnAl 258  
DB 181 TTTTGAGCGGTGAATCTGGAGCTTGTATTGAAAGCTACTCAAGTTAGATGTGAGTGT 240  
QY 258 aSerArgGlnGlyLeuGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAs 278  
DB 241 ATCGCGCGCAGGATATTGAAGATGCTGAAGAGAAAGCAGCTCAGCTTGTGGTGGCAGAGA 300  
QY 278 pSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAl 298  
DB 301 TACCACGAAGAGACTGTTTAAATATGATGAGATGAGGAGCACTGACCCCGAAGAAGAC 360  
QY 298 aglyProGluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMe 318  
DB 361 TCACGAGGAGCACCCTAACCATGCTCATCCACCGCAGAGCGCTCGAGCTCTGCT 420

QY 318 tSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGl 338  
DB 421 GTGCTGTGCTGTCTCTACATTGAGATGTCTGCCGTGTGCACGCTAAATTTGATAACA 480  
QY 338 yLysThrLysAspLeuTyrArgAspLeuLeuAsnIlePheAspLysLeuLeuLeuProTh 358  
DB 481 TAAACAAAGGATTATATACCGTGATCTGATATCCATCTTTTGACAAACTTCTGTGGCCAC 540  
QY 358 rHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPh 378  
DB 541 ACATGCCCTCTCCCATGTACAGTTCTTCATGTTTTTCTCTCAGCTTCAAGTTGGGATT 600  
QY 378 eAlaGluAlaPheLeuGluHisLeuTyrLysLysLeuGlnAspProSerAsnProAlaI 398  
DB 601 TGCAGAAGCATTTTGGAAACATCTCTGGAAAAGTTGCAGGATCCANATAACCCGCCAT 660  
QY 398 eIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIlePr 418  
DB 661 CATCAGGAGGCTGCTGCAAAATATATTGGGAGCTTTTGGCCAGAGCTAAATTTATTCC 720  
QY 418 oleuIleThrValLysSerCysLeuAspLeuLeuValAsnTyrLeuHisIleTyrLeuAs 438  
DB 721 TCTTATCACTGGAAGACATGCTGATCTCTTGNWTAACCTGGCTGCACATGTACCTTAC 780  
QY 438 nAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeu 452  
DB 781 TAACCAGGATTGGGAAC-AAGGCTTTTGTGACGTTGCACCTC 822  
BG502871 788 bp mRNA linear EST 27-MAR-2001  
602550430Fl NIH\_MGC\_61 Homo sapiens cDNA clone IMAGE:4658071 5',  
mRNA sequence.  
BG502871 GI:13464388  
EST.  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 788)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: CLONETECH Laboratories, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCMI451 row: 1 column: 08  
High quality sequence stop: 699.  
Location/Qualifiers  
1. .788  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4658071"  
/tissue\_type="embryonal carcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH\_MGC\_61"  
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1:  
SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc);  
Double-stranded cDNA was prepared from cell line RNA.  
and 3' adaptors were used in cloning as follows: 5',  
adaptor sequence: 5'-CAGCGCATATGACC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGCGCGGACATG-dr(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies

## FEATURES

### source

contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

## ORIGIN

```

Alignment Scores:
Pred. No.:      2 54e-134      Length:      788
Score:          1207.50      Matches:      245
Percent Similarity: 94.34%      Conservative: 5
Best Local Similarity: 92.45%      Mismatches: 10
Query Match:     35.48%      Indels:      6
DB:              12          Gaps:          1

US-09-932-678-2 (1-651) x BG502871 (1-788)

QY 302 ArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuVal 321
Db 2 CGGCTCGACACAGATGGTCATCTCTGAGCGCGCTGGACATCTCTGATGCTTTGGTT 61

QY 322 LeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspGlyLysThrLys 341
Db 62 TTGTCTTACNAGAGATGCTCTGATGATGATGGTAAGTTGATACGCCAAACAAAG 121

QY 342 AspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSer 361
Db 122 GATCTATATCGGACCTGATAAACAATTTTGACAACTCTCTGTGCCACCCATGCTCC 181

QY 362 CysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAla 381
Db 182 TGGCATGTACAGTTTTCATGTTTACCTCTGATGTTTCAAAATGGGATTCGCAGAGCA 241

QY 382 PheLeuGluHisLeuTyrLysLysLeuGlnAspProSerAsnProAlaIleAroGln 401
Db 242 TTTTGGAAACATCTCGAAAAAATTCAGAGCCCAAGTAATCTGCGCATCATCAGCAG 301

QY 402 AlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThr 421
Db 302 GCTGCTGGAATATATATGGAAGCTTINTGGCAAGAGCTAAATTTATCTCTTATCT 361

QY 422 ValLysSerCysLeuAspLeuLeuValAsnTyrPheHisIleTyrLeuAsnAsnGlnAsp 441
Db 362 GTAAATATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTATAACCCAGAT 421

QY 442 SerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGln 461
Db 422 TCGGGACAAAGGATCTCGCATGTTCCTCTCCATGGACCATTTTACTCAGCTGCCAA 481

QY 462 AlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGlu 481
Db 482 GCTGTGTCTACACCTTTGTTTGTAGACAAAGCAGCTTTTGAGCGGAAACCTGAAAGAA 541

QY 482 GlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPro 501
Db 542 GGTTCAGATATCTTCAGAGTCTGAAATTTTGACGGATAGTGTATGAGCCAGCTAAATCCC 601

QY 502 LeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsn-LysTyrG1 521
Db 602 CTGAAGATTGTCTG-CCCTCAGTGGTTAACTTTTGTGCAATCACAATTAAGTACCA 660

QY 521 nLeuValPheCysTyrThrIleIle-GluArgAsnAsnArgGln-MetLeuProVal-I1 540
Db 661 GCTGCTCTTCTGTATACCATCATTTGAGAGGAAACAATCGCAGAATGGTTGAGCTCCAT 720

QY 540 eArgSerThrAlaGlyClyAsp--SerValGlnIleCysThrAsnProLeuAspThrPh 559
Db 721 TGGAGTACCGCTGGAGGAGAACTCAGTGCAGCAATCTGGGCAAAACCCCGTGGAAACCT 780

QY 559 ePhePro 561
Db 781 TTTTCCC 787

```

RESULT 9

BF203684  
LOCUS  
DEFINITION

BF203684 996 bp mRNA linear EST 06-NOV-2000  
601866331F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4099398 5',  
mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS

BF203684  
BF203684.1 GI:11097270

SOURCE  
ORGANISM

Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 996)

## REFERENCE

NIH-MGC <http://mgc.nci.nih.gov/>.

## AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

## TITLE

Unpublished (1999)

## JOURNAL

Contact: Robert Strausberg, Ph.D.

## COMMENT

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LLCM967 row: n column: 07  
High quality sequence stop: 684.

## FEATURES

Location/Qualifiers

source

1..996

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone\_lib="NIH\_MGC\_17"

/tissue\_type="rhabdomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_17"

/note="Organ: muscle; Vector: pOTB7; Site: 1: EcoRI;

Site 2: XhoI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

## Alignment Scores:

Pred. No.: 2 88e-124 Length: 996  
Score: 1125.50 Matches: 238  
Percent Similarity: 88.24% Conservative: 2  
Best Local Similarity: 87.50% Mismatches: 26  
Query Match: 33.07% Indels: 9  
DB: 10 Gaps: 2

US-09-932-678-2 (1-651) x BF203684 (1-996)

QY 379 AlaGluAlaPheLeuGluHisLeuTyrLysLysLeuGlnAspProSerAsnProAlaIle 398  
Db 3 GCAGAGCGCATTTTGGACACATCTCTGAAAAAATG-CAGGACCCCAAGTAATCTCGCATC 61  
QY 399 IleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIlePro 418  
Db 62 ATCAGGACGCTGCTCGAAATATATATTGGAAGCTTTTCGGCAGAGCTAAATTTATTCCT 121  
QY 419 LeuIleThrValLysSerCysLeuAspLeuValAsnTyrLeuHisIleTyrLeuAsn 438  
Db 122 CTTATTTACTGTAATCATGCTAGATCTTGTGGTTAACTGGCTGCACATATACCTTAAT 181  
QY 439 AsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyPropheTyrSer 458  
Db 182 AACCCAGGATTCGGGAACAAAGGCATTCGCGATGTTGCTCTCCATGACCATTTTACTCA 241  
QY 459 AlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsn 478  
Db 242 GCCTGCCAAGCTGTGTCTACACCTTTGTTTTAGACACCAAGCAGCTTTTGAGCGGAAAC 301

```
Qy 479 LeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGln 498
Db 302 CTGAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAG 361
Qy 499 LeuAsnProLeuLysIleCysLeuProSerValValAsnPheAlaAlaIleThrAsn 518
Db 362 CTAATATCCCTGAGAGATTGCTGCCCTCAGTGGTTAACTTTTGTGCAATCACAAT 421
Qy 519 LysTyrGlnLeuValPheCysTyrThrIleLeuArgAsnAsnArgGlnMetLeuPro 538
Db 422 AAGTACCAAGCTCGCTCTCTACACCATCATTTGAGAGGAACAATCGCCAGATGTCGCA 481
Qy 539 ValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAsp-Th 558
Db 482 GTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTG-ACAAACCGCTGGACAAC 540
Qy 558 rPhePheProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTy 578
Db 541 CTTCTTCCCTTTGATCCCTGTGCTGAAGAGGTCAAGAAATTCATTGATCCTATTATA 600
Qy 578 rGlnValTrpGluAspMetSerAlaGluGlnLeuGlnPheLysLysProMetLysLys 598
Db 601 TCAGGTGTGGGAGACATGATGCTGAAGAGCTACAGAGTTCAGAGAAACCATGAAAA 660
Qy 598 sAspIleValGluAspGluAspAspPheLeuLysGlyGluValProGlnAsnAspTh 618
Db 661 GGACATAGTGAAGATGAAGATGATGC-ATTCTGAAGAGAGAAATGGCCAGATGGA----- 714
Qy 618 rValIleGlyIleThrProSerSerPheAspThrHisPheArgSerProSerSerVa 638
Db 715 ----TACGGGATGGGATTAGACAGCTCTTGGAACAGATTTGGAAGTCTACAAAAG----- 765
Qy 638 lGlySerProValLeuTyrMetGlnProSer 649
Db 766 -GGAGAGACCCACCAGATGATGACACCAAGT 798
RESULT 10
CAS111975 754 bp mRNA linear EST 15-NOV-2002
LOCUS
DEFINITION
UI-R-FJ0-cpx-i-24-0-UI-r1 UI-R-FJ0 Rattus norvegicus cDNA clone
UI-R-FJ0-cpx-i-24-0-UI 5', mRNA sequence.
CAS11975
ACCESSION
VERSION
KEYWORDS
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 754)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue procurement: Dr. James Lin, Universtiy of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 REVERSE
FEATURES
source
Location/Qualifiers
1..754
```

```
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-FJ0-cpx-i-24-0-UI"
/tissue_type="embryo"
/dev_stages="embryo"
/lab_host="DHI0B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-R-FJ0"
/notes="Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I;
UI-R-FJ0 is a cDNA library containing the following
tissue(s): rat embryo. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into p77T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dr)18 tail. The
sequence tag for this library is CATCTCTACT. This library
was created for the University of Iowa Program for Rat
Gene Discovery and Mapping (Val Sheffield, Bento Soares
and Tom Casavant)"
```

## ORIGIN

```
Alignment Scores:
Pred. No.: 8.66e-124 Length: 754
Score: 1120.00 Matches: 226
Percent Similarity: 51.39% Conservative: 15
Best Local Similarity: 48.19% Mismatches: 10
Query Match: 32.91% Indels: 218
DB: 14 Gaps: 1
US-09-932-678-2 (1-651) x CA5111975 (1-754)
Qy 139 AlaGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValPro 158
Db 1 GCACAGACTGTCTTCTCTAGACCATGCTCAGCATGATTGCTCTCTATTGTTGACCTCC 60
Qy 159 ArgValIleIleLysGluGlyAspValAspValSerAspSerAspGluAspAsn 178
Db 61 CGAGTTATTGTCAAGGAAGGTGGCATGATGTTTCAGATCTCAGAGTGAAGATGACAA 120
Qy 179 LeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleAlaArgTyrValPro 198
Db 121 CTCTCTGCAATTTTGACACATGTACAAAGCCTTGCAAAATAATAACAAGATATGCTCT 180
Qy 199 SerThrProTrpPheLeuMetProIleLeuValGluLysPhePheValArgLysSer 218
Db 181 TCGACACCATGTTTCTTAATGCCAATA----- 207
Qy 219 GluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThr 238
Db 207 ----- 207
Qy 239 LeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAla 258
Db 207 ----- 207
Qy 259 SerArgGlnGlyIleGluAspAlaGluThrAlaThrGlnThrCysGlyGlyThrAsp 278
Db 207 ----- 207
Qy 279 SerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAla 298
Db 207 ----- 207
Qy 299 GlyProGluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMet 318
Db 207 ----- 207
Qy 319 SerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGly 338
```

```
Db 207 ----- 207
QY 339 LysThrIysAspLeuTyrArgAspLeuLeuAsnIlePheAspLysLeuLeuProThr 358
Db 207 ----- 207
QY 359 HisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPhe 378
Db 207 ----- 207
QY 379 AlaGluAlaPheLeuGluHisLeuTrpLysLeuGlnAspProSerAsnProAlaIle 398
Db 207 ----- 207
QY 399 IleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIlePro 418
Db 207 ----- 207
QY 419 LeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsn 438
Db 208 ----- 208
QY 439 AsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSer 458
Db 247 AGCCAGGATTGGGAAACAAAGGCTTTTGTGATGTTGCATCCATGGACCATTTTATTC 306
QY 459 AlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsn 478
Db 307 GCTGTCAGAGCTGTGTTCTACATGTTGTTTGTAGACACAAAGCACTTTTGAGTGGCA 366
QY 479 LeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGln 498
Db 367 CTGAAGGAGGTCTACGGTATCTTCAGAGTCTAAATTTTGAACGATGTCGACCCAG 426
QY 499 LeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAlaIleThrAsn 518
Db 427 CTGAACCCACTGAAGATCTGCTGCCAGCGTGGTTAATTTCTTGCTGCCATAACAAT 486
QY 519 LysTyrGlnLeuValPheCysTyrThrIleGluArgAsnAsnArgGlnMetLeuPro 538
Db 487 AATATCCAGCTTGTGTTCTCTACATCATCGAGGAGGAAACATCGCCAGATGCTCCCT 546
QY 539 ValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThr 558
Db 547 GTTATTGGAAGTACAGCTGTGGAGACTCTGTGCAGACCTGCACCAACCATTTGGAC 606
QY 559 PhePheProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTyr 578
Db 607 TTCTTTCCCTTTGACCTTGTATGCTTAAGAGGTCAAAGAGTTCATCGATCCTGTCT 666
QY 579 GlnValTrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLys 598
Db 667 CAGCTTGGGAAGACATGAGTGTGAAGAGCTCAGGAGCTTANGAAACCCACTAGAAAG 726
QY 599 AspIleValGluAspGluAspAsp 607
Db 727 GAGCTAGTGGAGATGAAGACGATGAC 753

RESULT 11
CF738084
LOCUS
DEFINITION
IMAGE:30611903 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
```

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
cDNA Library preparation: Dr. James Lin University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mousefl.html  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5.  
Location/Qualifiers  
1. .769  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30611903"  
/tissue\_type="whole eye"  
/dev\_stage="embryo 12.5,13.5,14.5 dpc"  
/lab\_host="DH10B (ri phage resistant)"  
/clone\_lib="NIH BMAP HD0"  
/note="Organ: Eye; Vector: pYX- Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is TTTTGAAGT. This library was created for the University  
Iowa Brain Anatomy Project (BMAP): 'Gene discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH)."

ORIGIN  
Alignment Scores:  
Pred. No.: 2,36e-123 Length: 769  
Score: 1116.50 Matches: 212  
Percent Similarity: 90.62% Conservative: 20  
Best Local Similarity: 82.81% Mismatches: 23  
Query Match: 32.81% Indels: 1  
DB: 14 Gaps: 1  
US-09-932-678-2 (1-651) x CF738084 (1-769)  
QY 148 LeuSerMetIleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVal 167  
Db 3 CTCAGCATGATTCGGTCTCATTTTGTACCTCCCGAGTAATTTCTCAAGGAAGTGCATA 62  
QY 168 AspValSerAspSerAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHis 187  
Db 63 GATGTTTCAGATTCGTGACGACGAGATGACATCTCTTCGCAATTTTTCACATGTCAC 122  
QY 188 ArgAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProIle 207  
Db 123 AGAGCCCTTCAATAATAACAAGATATGTCACATGTCACACCATGTTCTTAATGCCAATA 182  
QY 208 LeuValGluLysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHis 227  
Db 183 CTGGTGGAGAAATTCCTCGTTTGTGAGGAAGTCCGAGAGAACATTTGGAATGTTATGTTCA 242  
QY 228 AsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIle 247  
Db 243 AACTTATTAAAGGATAAGTTATATTTCACCACTTTTCAGCGCGTGAATTTCTGGAGCTGTT 302  
QY 248 IleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGlu 267  
Db 303 ATTGAAAGCTACTCAAGTTAGATGTAGTGTATCGCGCAGGATATTGAAGATGCTGAA 362



```

VERSION BI561523.1 GI:15448837
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgs.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: Miklos Palkovits, M.D., Ph.D.
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11755 row: 1 column: 06
High quality sequence stop: 717.
FEATURES
    source
        Location/Qualifiers
            1..733
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5298581"
                /lab_host="DH10B"
                /clone_lib="NIH MGC 97"
                /notes="Organ: testis; Vector: pBluescriptR (modified
                pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
                (gtcgag); Oligo-dT primed using primer
                5'-TTTTTTTTTTTTTTVN-3', size-selected for average
                insert size 2.2 kb and normalized to R0F 5. This is a
                primary library enriched for full-length clones and
                constructed using the Cap-trapper method (Carninci, in
                preparation). Library constructed by M. Brownstein
                (NIMH/NHGRI). National Institutes of Health. Note: this is
                a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 5,36e-122 Length: 733
Score: 1105.00 Matches: 232
Percent Similarity: 97.48% Conservative: 0
Best Local Similarity: 97.48% Mismatches: 4
Query Match: 32.47% Indels: 5
DB: 12 Gaps: 0
US-09-932-678-2 (1-651) x BI561523 (1-733)
QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerAla 20
Db 23 ATGGCGGACCGCTGTTCACACGCGTTTCGGGAGATCGCGCGCTTCGCTCTGCA 82
QY 21 ValLysLeuGlyAlaSerArgThrGlyVIlleSerAsnMetArgAlaLeuGluAsnAsp 40
Db 83 GTTAAAGACCTGGCGCGCTGCAGACTGGGATTTCAATATGCTGATTAGAAATGAC 142
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluValLeu 60
Db 143 TTTTTCATTTCCCCCAAGAAAACACTGTTCGGTTTGGTGAACCTGTGACAGAACTTG 202
QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLysAsnGlnLeuLeuAsp 80
Db 203 CTGAAGTACAAAAGGTTGAAACAAATGACTTTTGAGTTGTGTAAGAACACGCTGTAGAT 262
QY 81 ProAspIleLysAspAspGlnIleAlaSerTrpLeuLeuGluPheArgSerIleMet 100
Db 263 CCAGACATAAAGATGACCAATCATCACTGGTGTCTAGAAATTCGTTCTTCTATCATG 322
QY 101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsn 120

```

```

Db 323 TACTTGACAAAGACATTTGAGCAACTTATCAGTATTATTAAGATTGCTTGGTTGAAT 382
QY 121 ArgSerGlnThrValValGluGluThrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140
Db 383 AGAAGTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCTTGGTAAATCTTGATCAGCACAG 442
QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProArgVal 160
Db 443 ACTGTTTTCCTCAGACCGGTCTCAGCATGATTGCTTCCCATTTTGTGCTTCCCGAGTG 502
QY 161 IleIleLysGluGlyAspValSerAspSerAspGluAspGluAspAsnLeuPro 180
Db 503 ATCATTAAGGAAGCGCATGTAGATGTTTCAGATTCGTAGTGAAGATGATATCTTCCT 562
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleAlaArgTyrValProSerThr 200
Db 563 GCAAAA-TTTGACACATGTCACAGAGCGCTTGCATAATAGCAAGATATGTACCATGCACA 621
QY 201 ProTrp-PheLeuMetProIleLeu-ValGluLysPheProPheValArgLysSerGluA 220
Db 622 CCGTGGGTTTCTCATGCCAATCTNGTGGAATAATTCATTTGTTGCG-AAATCAGAGA 680
QY 220 rgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPhe 236
Db 681 GAACACTGGATGTTAGCTTCACTTACTAAGGATTAG-GTATATTTC 729
RESULT 14
BG428305 1037 bp mRNA linear EST 14-MAR-2001
602498983F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4612605 5',
mRNA sequence.
ACCESSION BG428305
VERSION BG428305.1 GI:13334811
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgs.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1361 row: e column: 22
High quality sequence stop: 637.
FEATURES
    source
        Location/Qualifiers
            1..1037
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4612605"
                /lab_host="DH10B (T1 phage-resistant)"
                /clone_lib="NIH MGC 75"
                /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site 1:
                SfiI (ggcgctcgcc); Site 2: SfiI (ggccatagacc); 5' and
                3' adaptors were used in cloning as follows: 5' adaptor
                sequence: 5'-CACGCCCATATGGCC-3' and 3' adaptor sequence:
                5'-ATTCTAGAGCCGCGGCGGACATG-dn(30)BN-3' (where B = A,
                C, or G and N = A, C, G, or T). Average insert size 1.65
                kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
                by PCR. This library was enriched for full-length clones
                and was constructed by Clontech Laboratories (Palo Alto,
                CA). Note: this is a NIH_MGC Library."
ORIGIN

```



```
Alignment Scores:
Pred. No.: 3,19e-121 Length: 1037
Score: 1100.50 Matches: 252
Percent Similarity: 89.00% Conservative: 15
Best Local Similarity: 84.00% Mismatches: 19
Query Match: 32.34% Indels: 15
DB: 12 Gaps: 3

US-09-932-678-2 (1-651) x BG428305 (1-1037)
QY 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaSerSerSerAla 20
Db 22 ATGGCGGACCGCTGCTTACACCGGTTTGGCGGAGATGCGGCGCTTCGCTCTGCA 81
QY 21 ValLysLeuLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAsp 40
Db 82 GTTAAGAAGCTGGCGCGTCCGAGGACTGGGATTTCAATATATCGGTGCAATGAGATGAC 141
QY 41 PhePheAsnSerProProArgLysThrValArgPheGlyThrValThrGluValLeu 60
Db 142 TTTTTCATATTCCTCCCAAGAAACCTGTTCGGTTTGGTGGAACTGTGCACAGAAGTCTTG 201
QY 61 LeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80
Db 202 CTGAAGTACAAAGAGGTGAACAAATGACTTTGAGTTGTGAAGAACCGCTGTAGAT 261
QY 81 ProAspIleLysAspAspGlnIleLeuAsnTrpLeuLeuGluPheArgSerSerIleMet 100
Db 262 CCAGACATAAAGATGATGACCCAGATCATCAACTGGCTGTAGAAATTCGTTCTTCTATCATG 321
QY 101 TyrLeuThrLysAspPheGluGlnLeuLeuSerIleLeuLeuArgLeuProTrpLeuAsn 120
Db 322 TACTTGCACAAAAGACTTTGAGCAACTTATCAGTATATATATTAAGATTGCTTGGTTGAAT 381
QY 121 ArgSerGlnThrValValGluGluThrLeuAlaPheLeuGlyAsnLeuValSerAlaGln 140
Db 382 AGAAGTCAACAGTAGTGAAGAGTATTTGGCTTTTCTTGGTAATCTTGATACAGACAG 441
QY 141 ThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValProArg-Va 160
Db 442 ACTGTTTCTCTCAGACCGGTCTCAGCATGATTGCTTCCCATTTTGGCTTCCCGAAGT 501
QY 160 IleIleLysGluGlyAspValAspValSerAspAspGluAspAspAsnLeuPr 180
Db 502 GATCATTAAGGAAGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCC 561
QY 180 oAlaAsnPheAspThrCysHisArgAlaLeuGlnIleLeuAlaArgTyrValProSerTh 200
Db 562 TGCAAA-TTTGACACATGTACAGAGCCCTTGCAATATATAGCAAGATATGTACCATCGAC 620
QY 200 rPro-TripPheLeuMetProIleLeuVal-GluLysPheProPhe-ValArgLysSer- 218
Db 621 ACCGTTGGTTTCTCATGCCAACTACTGGGTGGCAACAATTTCCCATATGGTTCAGAAAACCTC 680
QY 219 --GluArgThrLeuGlu-CysTyr-ValHis-AsnLeuLeu-ArgIleSerValThrPhe 236
Db 681 CAGAGAGACACTGGAACCTGTACCGTTTCATGAACCTTACTTCAGGATCACGGGATATGTC 740
QY 237 ProThrLeu-ArgHis-GluIleLeuGluLeuIleLeuGluLysLeuLeuLysLeuAspV 256
Db 741 CCAAGCTTGGAGCGATGCAATCTGGAGCTTATTATTAAGAACTACTCAAGGTGGAGCTG 800
QY 256 aAlaAsnAlaSerArgGlnGlyIle---GluAspAla---GluGluThrAlaThrGlnThrC 274
Db 801 TGAATGGCTCCCGGGCGGGGTATTGGAAGACTGTGAGGAAGACGACGAAAGTACAACTG 860
QY 274 yGlyGlyThrAspSerThrGluGlyLeuPheAsnMet 286
Db 861 GTGCGGGAACAAATCCCGCAAGGGATTTGTCAGTTTG 898

RESULT 15
BE734354
```

LOCUS BE734354 757 bp mRNA linear EST 15-SEP-2000  
DEFINITION 601565596F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3840623 5', mRNA sequence.  
ACCESSION BE734354  
VERSION BE734354.1 GI:10148346  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 757)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LCM530 row: o column: 24  
High quality sequence stop: 693.  
location/Qualifiers  
1. .757  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3840623"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 21"  
/note="Organ: placenta; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAGG(G). Size-selected 500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN  
Alignment Scores:  
Pred. No.: 1,2e-120 Length: 757  
Score: 1094.00 Matches: 225  
Percent Similarity: 93.63% Conservative: 10  
Best Local Similarity: 89.64% Mismatches: 12  
Query Match: 32.15% Indels: 6  
DB: 10 Gaps: 0  
US-09-932-678-2 (1-651) x BE734354 (1-757)  
QY 181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleLeuAlaArgTyrValProSerThr 200  
Db 3 GCAAAATTTTGACACATGTCACAGAGCCTTGCAACTAATAGCAAGATATGTACATCGACA 62  
QY 201 ProTrp-PheLeuMetProIleLeuValGluLysPheProPheValArgLysSerGluAr 220  
Db 63 CGGTGATTTCTCATGCCAATACTGTTGGAAAAAATTTCCATTTGTTGAAAATCAGAGAG 122  
QY 220 gThrLeuGlu-CysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuA 240  
Db 123 AACACTGGACATGTTACCGTTCATAACTTACTAAGGATTAGTGTATATTTCACACCTTGA 182  
QY 240 rgHisGluIleLeuGluLeuIleGluLysLeuLeuLysLeuAspValAsnAlaSerA 260  
Db 183 GGCATGAAATTTCTGGAGCTTATTATTGAAAAAACAATCTCAAGTTGGATGTGATGATGCC 242  
QY 260 rgGlnGly-IleGluAspAlaGluThrAlaThrGlnThrCysGlyThrAspSer 279  
Db 243 GGCAGGGCTATTGAAGATGCTGAAGAAACAGCACTCAAACTTGTGGTGGACAGATTCC 302

```
QY 280 ThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLys-AlaG1 299
    |||||
Db 303 ACGGAAGGATTGTTTATATGGATGAAGATGAAGAACTGACATGNAACAAACGGCTGG 362
    |||||
QY 299 yProGluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSe 319
    |||||
Db 363 TCCTGAACGGCTCGACAGATGGTCATCCTGTAGCGAGCGCTGGACATCCTGATGTC 422
    |||||
QY 319 rLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsndglyLy 339
    |||||
Db 423 TTTGGGTTTGTCTCTACATGAAGGATGTCCTCTATGTAGATGGTAAGTTGATAACGSCAA 482
    |||||
QY 339 sThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHi 359
    |||||
Db 483 AACAAAGGATCTATATCGGACCTGATAACATCTTTGACAAACTCTGTAGCCCAACCA 542
    |||||
QY 359 sAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAl 379
    |||||
Db 543 TGCCTCTGCCATGTACAGTATTACATGTATTACCTCTGTAGTTTCAAAATCGGGATTGCG 602
    |||||
QY 379 aGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleI 399
    |||||
Db 603 AGAGGCATATTAGAACATCTCTGGAAAAAATTGCAGGCCCAAGTAATCTGCCATCAT 662
    |||||
QY 399 eArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLe 419
    |||||
Db 663 CAGGCAGGCTGTGGAAATTATATGGGAAGCTT-ATGGGAAGAGCCCTAATTATTCTCT 721
    |||||
QY 419 ulleThrValLysSerCysLeuAsp 427
    |||||
Db 722 ATTACTGTGTCCA-TCATGCCTAGAA 745
    |||||
```

Search completed: July 18, 2004, 12:07:14  
Job time : 4426 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 01:15:01 ; Search time 8111 Seconds  
(without alignments)  
11050.844 Million cell updates/sec

Title: US-09-932-678-1  
Perfect score: 2068  
Sequence: 1 acagaggctggtggaag.....tgacattgggattcccat 2068

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues 6940544

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hgc.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pi.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2068	100.0	2068	9	AF227156	AF227156 Homo sapi
2	2020	97.7	3756	9	HSA272050	AJ272050 Homo sapi
3	2018.4	97.6	2040	6	AX036047	AX036047 Sequence
4	2008	97.1	3709	9	BC036182	BC036182 Homo sapi
5	1211.6	58.6	135873	2	AC012547	AC012547 Homo sapi
6	1211.6	58.6	192815	9	AC013444	AC013444 Homo sapi
7	1179.8	57.1	3166	10	BC034110	BC034110 Mus muscu
8	894	43.2	1423	6	AX427479	AX427479 Sequence
9	567	27.4	2493	6	AX713860	AX713860 Sequence
10	567	27.4	2493	9	AK055742	AK055742 Homo sapi
11	506.4	24.5	1168	9	BC009198	BC009198 Homo sapi
12	478	23.1	2243	10	BC055781	BC055781 Mus muscu
13	420.8	20.3	1461	6	AX274907	AX274907 Sequence
14	420.8	20.3	2627	9	BC006441	BC006441 Homo sapi
15	420.8	20.3	3645	9	AK126166	AK126166 Homo sapi
16	420	20.3	437	6	BD073832	BD073832 5'EST of
17	292	14.1	341	6	BD072397	BD072397 Secreted
18	208.2	10.1	513	6	AR415449	AR415449 Sequence
19	208.2	10.1	513	6	BD111002	BD111002 EST and e
20	208	10.1	129473	2	AC040158	AC040158 Homo sapi
21	208	10.1	143783	9	AC139256	AC139256 Homo sapi
22	208	10.1	184471	9	AC136026	AC136026 Homo sapi
23	208	10.1	184891	9	AC092137	AC092137 Homo sapi
24	208	10.1	189770	2	AC139281	AC139281 Homo sapi
25	208	10.1	194670	2	AC141613	AC141613 Homo sapi
26	208	10.1	196687	2	AC136442	AC136442 Homo sapi
27	208	10.1	215222	2	AC141467	AC141467 Homo sapi
28	208	10.1	215974	2	AC141607	AC141607 Homo sapi
29	208	10.1	218593	2	AC141303	AC141303 Homo sapi
30	208	10.1	241278	9	AC097268	AC097268 Pan trogl
31	203.6	9.8	213	6	AX907303	AX907303 Sequence
32	203.6	9.8	213	6	BD042836	BD042836 Sequence
33	192	9.3	131839	9	AC112166	AC112166 Homo sapi
34	192	9.3	175691	2	AC007615	AC007615 Homo sapi
35	192	9.3	191496	2	AC009130	AC009130 Homo sapi
36	192	9.3	207997	9	AC009093	AC009093 Homo sapi
37	190.8	9.2	183586	9	AC092562	AC092562 Papio ham
38	189.8	9.2	218593	2	AC141303	AC141303 Homo sapi
39	188.8	9.1	173275	9	AC092375	AC092375 Homo sapi
40	188.8	9.1	179150	2	AC017077	AC017077 Homo sapi
41	188.8	9.1	191598	2	AC144362	AC144362 Gorilla g
42	188.8	9.1	202004	9	HUAF001549	AF001549 Human Chr
43	188.8	9.1	221030	2	AC145403	AC145403 Gorilla g
44	187.2	9.1	157631	2	AC142079	AC142079 Homo sapi
45	187.2	9.1	170613	9	AC106788	AC106788 Homo sapi

ALIGNMENTS

RESULT 1  
AF227156  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AF227156  
Homo sapiens RRN3 mRNA, complete cds.  
AF227156.1 GI:7670099  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 2068)  
Moorefield, B., Greene, E.A. and Reeder, R.H.  
RNA polymerase I transcription factor Rrn3 is functionally  
conserved between yeast and human

2068 bp mRNA linear PRI 29-APR-2000

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4724-4729 (2000)  
 MEDLINE 20243763  
 PUBMED 10758157  
 REFERENCE 2 (bases 1 to 2068)  
 AUTHORS Moorefield, B., Greene, E.A. and Reeder, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JAN-2000) Basic Sciences, Fred Hutchinson Cancer  
 Research Center, 1100 Fairview Avenue N., Seattle, WA 98109, USA  
 FEATURES  
 Location/Qualifiers  
 1..2068  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /map="16p12"  
 69..2024  
 /note="similar to yeast RNA polymerase I transcription  
 factor Rrn3"  
 factor Rrn3  
 /codon\_start=1  
 /product="RRN3"  
 /protein\_id="AAF66160.1"  
 /db\_xref="GI:7670100"  
 /translation="MAAPLLHRLPCDAASSAVKLGASRTGISNMRLNLEDFNS  
 PRKTVRFQGTVEVLLKYKGETNDFELLNQLLDPOIKDDQIIINWLLFSSIMYL  
 TKDFQLISIIIRLPLWNSQTVREYLAFLGNLSAOTVFLRPLCLMLASHFVPRV  
 IIKEDVDVSDDEDDNLPANFDIHRALQIIARYVSTPWFELPILVEKEFPFKRS  
 ERLCEYVHNLRISYIFPTLRHIELELLEIKLKDVLNRSQGHEDAEATATQCGG  
 TDSTGLFNMDRETEHETKAPERLDQVHVAERLDILMSLVLSYNKOVYVDGK  
 VNGKTDILYRLINIIFDKLLPHTASCHQVFMFVLCFSFKLGEAEFLHLEHKKLQD  
 PNPATIROAGNYIGSFARAKFIPLITVKSCDLLVNWHLIYLNQDSGTFKAFCDV  
 ALHGPYSACQVFTFVFRHKLSSGNLKEGLVYLSNFERIIVMSQNLPLKICLPS  
 VYNFPAATNKQVFCYTLIERNSQMLPVIRSTAGDSVOICTNPLDTEFPFPCV  
 LKRSKFIPIQVWEDMSAELEQFKPKKKDIIVDEDDDFLKSEVPQNDIVGILTP  
 SSFDTHFRSPSSVSGPPVLYMQPSPL"

CDS

ORIGIN  
 Query Match 100.0%; Score 2068; DB 9; Length 2068;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2068; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCTCAGGCGCAGCGGTCCGTTAGT 60  
 DB 1 ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCTCAGGCGCAGCGGTCCGTTAGT 60  
 QY 61 TCGGCCCAATGCGGCGCACCGCTGCTTACACGCGTTTTCGGGAGAGTGGCGCGTTCGT 120  
 DB 61 TCGGCCCAATGCGGCGCACCGCTGCTTACACGCGTTTTCGGGAGAGTGGCGCGTTCGT 120  
 QY 121 CCTCTGCAGTTAAGAGCTGGGCGCGTTCGAGGACTGGGATTTCAAATATCGGTGCATTAG 180  
 DB 121 CCTCTGCAGTTAAGAGCTGGGCGCGTTCGAGGACTGGGATTTCAAATATCGGTGCATTAG 180  
 QY 181 AGAATGACTTTTCAATTCTCCCAAGAAAACTGTTTCGGTTTGGTGGAACTGTGCACAG 240  
 DB 181 AGAATGACTTTTCAATTCTCCCAAGAAAACTGTTTCGGTTTGGTGGAACTGTGCACAG 240  
 QY 241 AAGTCTTGTGAGTACAAAAGGGTGAACAATGACTTTGAGTTGTGTAAGAACCCAGC 300  
 DB 241 AAGTCTTGTGAGTACAAAAGGGTGAACAATGACTTTGAGTTGTGTAAGAACCCAGC 300  
 QY 301 TGTTAGATCCAGACATAAAGGATGACCAAGATCATCAACTGGCTGCTAGAAATCCGTTCTT 360  
 DB 301 TGTTAGATCCAGACATAAAGGATGACCAAGATCATCAACTGGCTGCTAGAAATCCGTTCTT 360  
 QY 361 CTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATATATTAAGATTGCTT 420  
 DB 361 CTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATATATTAAGATTGCTT 420  
 QY 421 GGTGTAATAGAAGTCAAAACAGTACTGGAAGATTTTGGCTTTTCTTGGTAACTTGTAT 480  
 DB 421 GGTGTAATAGAAGTCAAAACAGTACTGGAAGATTTTGGCTTTTCTTGGTAACTTGTAT 480  
 QY 481 CAGCACAGACTGTTTCTCCTCAGACCGGTGCTCAGCATGATGTGCTTCCCAATTTTGTGCTC 540

DB 481 CAGCACAGACTGTTTCTCCTCAGACCGGTGCTCAGCATGATGCTTCCCAATTTTGTGCTC 540  
 QY 541 CCCGAGTGATCATTAAGGAAGCGGATAGATGTTTTCAGATCTCAGATGATGAAGATGATA 600  
 DB 541 CCCGAGTGATCATTAAGGAAGCGGATAGATGTTTTCAGATCTCAGATGATGAAGATGATA 600  
 QY 601 ATCTCTCTGCAAAATTTTGACACATGTCACAGAGCCTTGCAAAATAATAGCAAGATATGAC 660  
 DB 601 ATCTCTCTGCAAAATTTTGACACATGTCACAGAGCCTTGCAAAATAATAGCAAGATATGAC 660  
 QY 661 CATCGACACCGGTGTTTCTCATGCCAATACTGSGTGGAAAAATTTCCATTTGTTCCAAAT 720  
 DB 661 CATCGACACCGGTGTTTCTCATGCCAATACTGSGTGGAAAAATTTCCATTTGTTCCAAAT 720  
 QY 721 CAGAGAGAACTGGAATGTTAGCTTACTAAGGATGATGATATTTCCAA 780  
 DB 721 CAGAGAGAACTGGAATGTTAGCTTACTAAGGATGATGATATTTCCAA 780  
 QY 781 CCTTCAGGATGAAATTTCTGAGCTTATTATTGAAAAAATCTCAAGTTGGAATGTGAATG 840  
 DB 781 CCTTCAGGATGAAATTTCTGAGCTTATTATTGAAAAAATCTCAAGTTGGAATGTGAATG 840  
 QY 841 CATCCCGCAGGCTTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGCTGGACAG 900  
 DB 841 CATCCCGCAGGCTTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGCTGGACAG 900  
 QY 901 ATTCCAGGAAGATGTTTAATATGGAATGAAGAACTGAACATGAACAAAGG 960  
 DB 901 ATTCCAGGAAGATGTTTAATATGGAATGAAGAACTGAACATGAACAAAGG 960  
 QY 961 CTGCTCTCAAGCGCTCGACAGATGCTGATCTGATAGCGGCTGGACATCCTGA 1020  
 DB 961 CTGCTCTCAAGCGCTCGACAGATGCTGATCTGATAGCGGCTGGACATCCTGA 1020  
 QY 1021 TGTCTTTGTTTGTCTACATGAAGGATGCTGCTATGATAGTGTAGGTTGATGAACG 1080  
 DB 1021 TGTCTTTGTTTGTCTACATGAAGGATGCTGCTATGATAGTGTAGGTTGATGAACG 1080  
 QY 1081 GCAAAACAAAGATCTATATCGGACCTGATTAACATCTTTCGACAACTCCTGTCGCCA 1140  
 DB 1081 GCAAAACAAAGATCTATATCGGACCTGATTAACATCTTTCGACAACTCCTGTCGCCA 1140  
 QY 1141 CCATGCTCTCTGCAATGACAGTCTTTCATGTTTTCCTGCTGCTGCTGCTGCTGCTGCT 1200  
 DB 1141 CCATGCTCTCTGCAATGACAGTCTTTCATGTTTTCCTGCTGCTGCTGCTGCTGCTGCT 1200  
 QY 1201 TCGCAGAGCAATTTTGGAAACATCTCTGAAAAAATTTGAGGACCCCAAGTAATCTCTGCCA 1260  
 DB 1201 TCGCAGAGCAATTTTGGAAACATCTCTGAAAAAATTTGAGGACCCCAAGTAATCTCTGCCA 1260  
 QY 1261 TCATCAGGCGAGCTGCTGGAATTTATTTGGAGCTTTTGGCAAGCTTAAATTTTATTC 1320  
 DB 1261 TCATCAGGCGAGCTGCTGGAATTTATTTGGAGCTTTTGGCAAGCTTAAATTTTATTC 1320  
 QY 1321 CTCTTATTAAGTAAATCATGCTAGATCTTTTGGTTAACTGCTGCTGCTGCTGCTGCTGCT 1380  
 DB 1321 CTCTTATTAAGTAAATCATGCTAGATCTTTTGGTTAACTGCTGCTGCTGCTGCTGCTGCT 1380  
 QY 1381 ATAAACAGGATTTGGGAAACAAAGGCAATTCGATGTTGCTCTCCATGACCAATTTTACT 1440  
 DB 1381 ATAAACAGGATTTGGGAAACAAAGGCAATTCGATGTTGCTCTCCATGACCAATTTTACT 1440  
 QY 1441 CAGCCTGCCAAGCTGTTCTACACCTTTGTTTGTAGACAAAGCAGCTTTTTCAGCGGAA 1500  
 DB 1441 CAGCCTGCCAAGCTGTTCTACACCTTTGTTTGTAGACAAAGCAGCTTTTTCAGCGGAA 1500  
 QY 1501 ACCTGAAAGAGGTTTTCAGTATCTTTCAGAGTCTGAATTTTTCAGCGGATGATGATGAGCC 1560  
 DB 1501 ACCTGAAAGAGGTTTTCAGTATCTTTCAGAGTCTGAATTTTTCAGCGGATGATGATGAGCC 1560  
 QY 1561 AGCTAAATCCCTGAAAGATTTGCTCCCTCAGTGGTAACTTTTTCGCTGCAATCAAA 1620

```
Db 1561 AGCTAAATCCCTCGAAGATTGCTGCCCTCAGTGTAACTTTTGTGCAATCACAA 1620
QY 1621 ATAAGTACAGCTCGTCTTCTGTGCTACACCATCATTCAGAGGAACAATCGCAGATCTGC 1680
Db 1621 ATAAGTACAGCTCGTCTTCTGTGCTACACCATCATTCAGAGGAACAATCGCAGATCTGC 1680
QY 1681 CAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACA 1740
Db 1681 CAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACA 1740
QY 1741 CCTTCTCCCTTGTGATCCCTGTGTCTGAAGAGGTCAAAGAAATTCATTGATCTTATTT 1800
Db 1741 CCTTCTCCCTTGTGATCCCTGTGTCTGAAGAGGTCAAAGAAATTCATTGATCTTATTT 1800
QY 1801 ATCAGTATGGGAAGACATAGTGTCTGAAGAGTACAGAGGTTCAAAGAAACCCATGAAAA 1860
Db 1801 ATCAGTATGGGAAGACATAGTGTCTGAAGAGTACAGAGGTTCAAAGAAACCCATGAAAA 1860
QY 1861 AGGACATAGTGAAGATGAAGATGATGACTTCTGAAGGCGAAGTGCCCGCAGATGATA 1920
Db 1861 AGGACATAGTGAAGATGAAGATGATGACTTCTGAAGGCGAAGTGCCCGCAGATGATA 1920
QY 1921 CGTGATTTGGGATCACACCAAGCTCCTTTGACACGCAATTCGGAAGTCTCTCAAGTAGTG 1980
Db 1921 CGTGATTTGGGATCACACCAAGCTCCTTTGACACGCAATTCGGAAGTCTCTCAAGTAGTG 1980
QY 1981 TGGGCTCCCCACCGTGTGTACATGCAACCCAGTCCCTCTGACGGCAGAAATTTGTGA 2040
Db 1981 TGGGCTCCCCACCGTGTGTACATGCAACCCAGTCCCTCTGACGGCAGAAATTTGTGA 2040
QY 2041 CTGAGATGTGACATTTGGGATTTCCCAT 2068
Db 2041 CTGAGATGTGACATTTGGGATTTCCCAT 2068

RESULT 2
HSA272050
LOCUS HSA272050 3756 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens mRNA for transcription initiation factor IA protein
(TIF-IA gene).
ACCESSION AJ272050
VERSION AJ272050.1 GI:10046713
KEYWORDS TIF-IA; transcription initiation factor.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Bodem,J., Dobrev,G., Hoffmann-Rohrer,U., Iben,S., Zentgraf,H.,
Delius,H., Vingron,M. and Grummt,I.
TIF-IA, the factor mediating growth-dependent control of ribosomal
RNA synthesis, is the mammalian homolog of yeast Rn3p
EMBO Rep. 1 (2), 171-175 (2000)
21163499
PUBMED 11265758
REFERENCE 2 (bases 1 to 3756)
Bodem,J.J.
Direct Submission
Submitted (07-FEB-2000) Bodem J.J., Molekulare Biologie der Zelle
II, Deutsches Krebsforschungszentrum, INF.280, 69120 Heidelberg,
GERMANY
FEATURES
source Location/Qualifiers
1..3756
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..3756
/gene="TIF-IA"
23..1978
/gene="TIF-IA"
/function="transcription initiation factor"
/codon_start=1
/evidence=experimental
```

```
/product="transcription initiation factor IA protein"
/protein_id="CAC07955.1"
/db_xref="GI:10046714"
/db_xref="GOA:Q9H4F0"
/db_xref="SPTREMBL:Q9H4F0"
/translation="MAAPLLHTRLPGDAAASSAVKKLGASRTGISMRALENDFFNS
PRKTVRFQGTVEVLKYKGETNDFELLKQLLPDIDKDDQIINWLEFFSSIMVL
TKDFPOLISIIILRLPWLNRSTQVBEYLAFLGNLYSAQTVFLRPLCLSMIAHFVPRV
IIEKGDVDSDDDDNLNPNFDTCHRALQIILIARYVSTPWFLLVEKFPFVRKS
RETLCYVHNLIRISYIFTLRHELELEIIEKLLKLDVNASQROGEDAETATQTCGG
DUSTGLNMDDEETETKAGPERLDQVHPVAERLDILMSLVSYMKDVCYVDGK
VNDGKOLYRDLINI FDKLLPHTASCHVQFMFYLCFSKLGFAEAFLEHLWKQLQD
PSNPATIRQAAGNYIGSLRAKAFIPLITVKSLDLLVNLHLIYVNNQDSGTKAFCDV
ALHGPYSACOAIFYFVFRHKQLLSGNLKEGLVQLSINFERIYVNNQSLNPLKICLPS
VNVFAATITNKVLVECYTILERNRQMLPVIRSTAGGDSVQICNPLDTFFPFDDCV
LKRSKFFDPIYQWEDMSABELQFKKPMKKDIIVDEDDDDFLKGEVPPQNDTVIGTTP
SSFDHFRSPSSSVGSPVLYNQPSPL"
1979..3756
/gene="TIF-IA"
/evidence=experimental

3'UTR
ORIGIN
Query Match 97.7%; Score 2020; DB 9; Length 3756;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2020; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 47 GCGGTCGCGTTAGTTCGGCCCAATCGCGCACCGCTGCTTCACACGCTTCCCGGGAGA 106
Db 1 GCGGTCGCGTTAGTTCGGCCCAATCGCGCACCGCTGCTTCACACGCTTCCCGGGAGA 60
QY 107 TCGGCGCGCTTCGTCCTCTGCAGTTAAGAAGCTGCGGCGCTCGAGGACTGGGATTTCAA 166
Db 61 TCGGCGCGCTTCGTCCTCTGCAGTTAAGAAGCTGCGGCGCTCGAGGACTGGGATTTCAA 120
QY 167 TATGCGTGCAATTAGAGAATGACTTTTCAATTTCTCCCCCAAGAAAAACTGTTTCGGTTGG 226
Db 121 TATGCGTGCAATTAGAGAATGACTTTTCAATTTCTCCCCCAAGAAAAACTGTTTCGGTTGG 180
QY 227 TGGAACTGTGACAGAGTCTTGTGAGTACAAAAGAGGTGAAACAATGACTTTGAGTT 286
Db 181 TGGAACTGTGACAGAGTCTTGTGAGTACAAAAGAGGTGAAACAATGACTTTGAGTT 240
QY 287 GTTGAAGAACCGAGCTGTTTAGATCCAGACATAAAGGATGACAGATCATCAACTGCTGCT 346
Db 241 GTTGAAGAACCGAGCTGTTTAGATCCAGACATAAAGGATGACAGATCATCAACTGCTGCT 300
QY 347 AGAATTCGGTCTTCTATCATGTACTTGAACAAAAGACTTTGAGCAACTTATCAGTATTAT 406
Db 301 AGAATTCGGTCTTCTATCATGTACTTGAACAAAAGACTTTGAGCAACTTATCAGTATTAT 360
QY 407 ATTAAGATTGCTTGGTTGAATAGAGTCAACAGTAGTAGTGAAGAGTATTTGGCTTTTCT 466
Db 361 ATTAAGATTGCTTGGTTGAATAGAGTCAACAGTAGTAGTGAAGAGTATTTGGCTTTTCT 420
QY 467 TGGTAATCTTGTATCAGCAGACTGTTTTCTCCTCAGACCGGTCTCAGCATGATTGCTTC 526
Db 421 TGGTAATCTTGTATCAGCAGACTGTTTTCTCCTCAGACCGGTCTCAGCATGATTGCTTC 480
QY 527 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGCGAGTAGTAGATTGTTTCAAGATTCTGA 586
Db 481 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGCGAGTAGTAGATTGTTTCAAGATTCTGA 540
QY 587 TGATGAAGATGATAATCTTCTGCAAAATTTTGACATGTACAGAGCCCTTGCAAAATAT 646
Db 541 TGATGAAGATGATAATCTTCTGCAAAATTTTGACATGTACAGAGCCCTTGCAAAATAT 600
QY 647 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATCTGCTGGAATAATTTCC 706
Db 601 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATCTGCTGGAATAATTTCC 660
QY 707 ATTGTTTCGAAATTCAGAGAAACATCTGGAATGTTACGTTTATAACTTACTAAGGATTAG 766
Db 661 ATTGTTTCGAAATTCAGAGAAACATCTGGAATGTTACGTTTATAACTTACTAAGGATTAG 720
```

QY	767	TGTAATATTTTCCAACTTGAGGATGAATTTCTGGAGCTTATTAATGAAAACTACTCAA	826
Db	721	TGTATATTTTCCAACTTGAGGATGAATTTCTGGAGCTTATTAATGAAAACTACTCAA	780
QY	827	GTTGATGTAATGATCATCCGACGAGTATTCAGATGCTGAAGAAACAGCAACTCAAC	886
Db	781	GTTGATGTAATGATCATCCGACGAGTATTCAGATGCTGAAGAAACAGCAACTCAAC	840
QY	887	TTGTGCTGGGACAGATTTCCACGGAAGCATTTGTTTAATATGGATGAAGATGAAGAACTGA	946
Db	841	TTGTGCTGGGACAGATTTCCACGGAAGCATTTGTTTAATATGGATGAAGATGAAGAACTGA	900
QY	947	ACATGAACAAAGGCTGCTTGAACGCTCGACGATGGTGGATCTGTGATCCGAGCG	1006
Db	901	ACATGAACAAAGGCTGCTTGAACGCTCGACGATGGTGGATCTGTGATCCGAGCG	960
QY	1007	CTTGACATCTGATCTCTTTGTTTGTCTACATGAAGGATGTCTGTATGTAGATCG	1066
Db	961	CTTGACATCTGATCTCTTTGTTTGTCTACATGAAGGATGTCTGTATGTAGATCG	1020
QY	1067	TAAGGTTGATAACGGCAAAACAAAGGATCTATATCGGACCTGATATAACATCTTTGACAA	1126
Db	1021	TAAGGTTGATAACGGCAAAACAAAGGATCTATATCGGACCTGATATAACATCTTTGACAA	1080
QY	1127	ACTCCTGTTGCCACCATGCTCTCCATGATGACAGTCTTTCATCTTTTACCTCTGTAG	1186
Db	1081	ACTCCTGTTGCCACCATGCTCTCCATGATGACAGTCTTTCATCTTTTACCTCTGTAG	1140
QY	1187	TTTCAAAATTTGGGATTCGACAGGCAATTTTGGAAACATCTCTGAAAAAATTCGAGACCC	1246
Db	1141	TTTCAAAATTTGGGATTCGACAGGCAATTTTGGAAACATCTCTGAAAAAATTCGAGACCC	1200
QY	1247	AAGTAATCTGCCATCATCAGGCAGGCTGCTGGAATATATTTGAAAGCTTTTGGCAAG	1306
Db	1201	AAGTAATCTGCCATCATCAGGCAGGCTGCTGGAATATATTTGAAAGCTTTTGGCAAG	1260
QY	1307	AGCTAAATTTTCTCTTATTAATGATTAATCATGCTAGATCTTTTGGTTAACTGGCT	1366
Db	1261	AGCTAAATTTTCTCTTATTAATGATTAATCATGCTAGATCTTTTGGTTAACTGGCT	1320
QY	1367	GCACATATCTTTAAATACCAAGGATTCGGAAACAAAGGCAATCTGCGATCTTCTCTCCA	1426
Db	1321	GCACATATCTTTAAATACCAAGGATTCGGAAACAAAGGCAATCTGCGATCTTCTCTCCA	1380
QY	1427	TGGAACATTTTACTCAGCCTGCCAAGCTGTGTTCTACACTTTGTTTTTGAACAAGCA	1486
Db	1381	TGGAACATTTTACTCAGCCTGCCAAGCTGTGTTCTACACTTTGTTTTTGAACAAGCA	1440
QY	1487	GCTTTTGAGCGGAAACCTGAAAGAGTTTGGAGTATCTTCAGAGTCTGAATTTTGAGCG	1546
Db	1441	GCTTTTGAGCGGAAACCTGAAAGAGTTTGGAGTATCTTCAGAGTCTGAATTTTGAGCG	1500
QY	1547	GATAGTGATGAGCGAGCTAAATCCCTTGAAGATTTGCTGCCCTCAGTGGTTAACTTTT	1606
Db	1501	GATAGTGATGAGCGAGCTAAATCCCTTGAAGATTTGCTGCCCTCAGTGGTTAACTTTT	1560
QY	1607	TGCTGCAATCACAAATAGTACAGCTCTCTTCTGTACACCATCTTTCAGAGGAACAA	1666
Db	1561	TGCTGCAATCACAAATAGTACAGCTCTCTTCTGTACACCATCTTTCAGAGGAACAA	1620
QY	1667	TCGCCGATCTGCGAGTATAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCAC	1726
Db	1621	TCGCCGATCTGCGAGTATAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCAC	1680
QY	1727	AAACCCGCTGGACACCTTTCTCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAAT	1786
Db	1681	AAACCCGCTGGACACCTTTCTCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAAT	1740
QY	1787	CATTGATCTTATTTATCAGGTATGGGAAGACATGAGTCTGGAAGCTACAGGATTCAA	1846
Db	1741	CATTGATCTTATTTATCAGGTATGGGAAGACATGAGTCTGGAAGCTACAGGATTCAA	1800
QY	1847	GAACCCCATGAAAAAGGACATAGTGAAGATGAAGATGATGATCTTCTGAAAGCGGAAGT	1906
Db	1801	GAACCCCATGAAAAAGGACATAGTGAAGATGAAGATGATGATCTTCTGAAAGCGGAAGT	1860
QY	1907	GCCCCAGAAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCAATTCGGAAG	1966
Db	1861	GCCCCAGAAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCAATTCGGAAG	1920
QY	1967	TCCTTCAAGTAGTGTGGGCTCCCCACCGTGTTGTACATGCAACCCAGTCCCTCTCGAG	2026
Db	1921	TCCTTCAAGTAGTGTGGGCTCCCCACCGTGTTGTACATGCAACCCAGTCCCTCTCGAG	1980
QY	2027	GCAGAAATTTGTGACTGAGATGTCATTTGGGATTTCCCAT 2068	
Db	1981	GCAGAAATTTGTGACTGAGATGTCATTTGGGATTTCCCAT 2022	
RESULT 3			
AX036047			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
CDS			
Location/Qualifiers			
1..2040			
/organism="Homo sapiens"			
/mol_type="unassigned DNA"			
/db_xref="taxon:9606"			
40..1995			
/note="unnamed protein product"			
/codon_start=1			
/protein_id="CAC16268.1"			
/db_xref="GI:1191587"			
/db_xref="REMTREMBL:CAC16268"			
/translation="MAAPLHTLPGDAASAVKKGASRTGISNMRALDNFNS			
PRKTVRFPGTVEVLTKYKGETNDFELLKNQLLDPDKDDIINWLEFRSSIMYL			
TKDFEQLISIIILPWLNSQTVVEYLAPLGNLVAQTVFURPCLSMIASHFVPRV			
IIEGDDVSDSDDEDDNLPAFNDTCHRALQIARYVPSTPWFMLPILVEKFPFVRKS			
ERTLECQVHLKLRISVYPTLRHLEILELIEKLKLDVNASQGIEDABETATQCGG			
TDSLEGLFNWDEDETEHETKAGPERLDQWHPVAERLDILMSLVSMKDVQCYDGK			
VDSKTKDLYRDLINIFDKLLLPHTASCHVQPFMFVLCSPKLGFAFAFLEHLKFAFCDV			
PNSPALIRAGNYIGSLFARAKFPIELIVKSCDLLLVNHLIYLNQDSGTAKFCDV			
ALHPFPYSQAVFYAFVFRKOLLGSLNGLQYLOSLNFRIVMSQLNPLKILCLPS			
VNFVFAATIKYQLVFCYTIERNRNQMLPVIRSTAGSDSVQICTNPLDTFFPDPVCV			
LKRKKFIDPIYQVWEDMSAELEKFKPKMKDIVERDEDDDFLKGVEPQNDTVIGTTP			
SSPDTHFRSPSSSVGSPVLYMQPSPL"			
ORIGIN			
Query Match 97.6%; Score 2018.4; DB 6; Length 2040;			
Best Local Similarity 99.9%; Pred. No. 0;			
Matches 2019; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	47	GCAGTGCAGTTCAGTTCGGCCCAATGCGGCACCGCTGCTTTCACACGGCTTTGCGGAGAG	106
Db	18	GCAGTGCAGTTCAGTTCGGCCCAATGCGGCACCGCTGCTTTCACACGGCTTTGCGGAGAG	77
QY	107	TGCGGCGCGCTTCGCTCTCTCAGTTAAGAACTGGGCGCGCTCGAGGACTCGGATTTCAA	166
Db	78	TGCGGCGCGCTTCGCTCTCTCAGTTAAGAACTGGGCGCGCTCGAGGACTCGGATTTCAA	137
QY	167	TATCGTGCATATAGAGATGACTTTTCAATTTCTCCCCCAAGAAAACCTGTCGTTGG	226

Db 138 TATGCGTCATTAGAGAAATGACTTTTCAATTCTCCCCCAAGAAAACCTGTTCCGGTTGG 197  
QY 227 TGGAACTGTGACAGAACTCTTCTGAACTPACAAAAGGGTGAACCAAAATGACTTTGAGTT 286  
Db 198 TGGAACTGTGACAGAACTCTTCTGAACTPACAAAAGGGTGAACCAAAATGACTTTGAGTT 257  
QY 287 GTTGAAGAACCAAGCTGTGTAGATCCAGACATTAAGAGTACCAAGATCAATCAACTGGTGTCT 346  
Db 258 GTTGAAGAACCAAGCTGTGTAGATCCAGACATTAAGAGTACCAAGATCAATCAACTGGTGTCT 317  
QY 347 AGAATTCGGTCTTCTTATCATGTAATTGACAAAAGACTTTTGAAGCACTTATCAGATTATAT 406  
Db 318 AGAATTCGGTCTTCTTATCATGTAATTGACAAAAGACTTTTGAAGCACTTATCAGATTATAT 377  
QY 407 ATTAAGATTCGGTCTTCTTGAATPAGAACTCAACAGTGTGGAAGAGTATTTGGCTTTTCT 466  
Db 378 ATTAAGATTCGGTCTTCTTGAATPAGAACTCAACAGTGTGGAAGAGTATTTGGCTTTTCT 437  
QY 467 TGGTAATCTGTATCAGCACAGACTGTTTCTCAGACGGTGTCTCAGATGATTTGCTTC 526  
Db 438 TGGTAATCTGTATCAGCACAGACTGTTTCTCAGACGGTGTCTCAGATGATTTGCTTC 497  
QY 527 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAAGATTCGA 586  
Db 498 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAAGATTCGA 557  
QY 587 TGATGAAGATGATTAATCTTCTGCAAAATTTTGACACATGTCTACAGAGCTTTGCAAAATAT 646  
Db 558 TGATGAAGATGATTAATCTTCTGCAAAATTTTGACACATGTCTACAGAGCTTTGCAAAATAT 617  
QY 647 AGCAAGATGTATPACCATCGACACCGTGTCTCATGCCAATCTGTTGGTGAAGAAATTTCC 706  
Db 618 AGCAAGATGTATPACCATCGACACCGTGTCTCATGCCAATCTGTTGGTGAAGAAATTTCC 677  
QY 707 ATTTGTTCCAAAATCAGAGAGAAACAATGGAATGTACGTTTCAATCTTACTAAGATTAG 766  
Db 678 ATTTGTTCCAAAATCAGAGAGAAACAATGGAATGTACGTTTCAATCTTACTAAGATTAG 737  
QY 767 TGTATATTTTCCAACTTGAGGATGAAATCTGGAGCTTATTTGAAAACTACTCAA 826  
Db 738 TGTATATTTTCCAACTTGAGGATGAAATCTGGAGCTTATTTGAAAACTACTCAA 797  
QY 827 GTTCGATGTGATCATCCCGCGAGGATTTGAAGATCTGAGAGAAACAGCACTCAAC 886  
Db 798 GTTGAATGTGAATCATCCCGCGAGGATTTGAAGATCTGAGAGAAACAGCACTCAAC 857  
QY 887 TTGTGGTGGGACAGATTCACCGAAGGATTTGTTAATATGATGAAGATGAAGAACTGA 946  
Db 858 TTGTGGTGGGACAGATTCACCGAAGGATTTGTTAATATGATGAAGATGAAGAACTGA 917  
QY 947 ACATGAACAAAGGCTGTCTGAACCGCTCGACAGATGTGCATCTCTGAGCGGAGCG 1006  
Db 918 ACATGAACAAAGGCTGTCTGAACCGCTCGACAGATGTGCATCTCTGAGCGGAGCG 977  
QY 1007 CCTGCATCTCTGATGTCTTTGTTTGTCTCTACATGAGGATCTCTGATGTAGATCG 1066  
Db 978 CTTGCACATCTCTGATGTCTTTGTTTGTCTCTACATGAGGATCTCTGATGTAGATCG 1037  
QY 1067 TAAGGTTGATAACGGCAAAAACAAAGGATCTATATCGGACCTGATAAATCTTTTGACAA 1126  
Db 1038 TAAGGTTGATAACGGCAAAAACAAAGGATCTATATCGGACCTGATAAATCTTTTGACAA 1097  
QY 1127 ACTCTGTGTGCCACCCANGCTCTCTGCATGTACAGTTTTCATGTTTACCTCTGTAG 1186  
Db 1098 ACTCTGTGTGCCACCCANGCTCTCTGCATGTACAGTTTTCATGTTTACCTCTGTAG 1157  
QY 1187 TTTCAAAATTTGGATTCGAGAGGATTTTTCGAACATCTCTGGAAGAAATTTGAGGACCC 1246  
Db 1158 TTTCAAAATTTGGATTCGAGAGGATTTTTCGAACATCTCTGGAAGAAATTTGAGGACCC 1217  
QY 1247 AAGTAATCTCTGCCATCATCAGCGAGGCTGTGGAAATTTATATTGGAAGCTTTTGGCAAG 1306  
Db 1218 AAGTAATCTCTGCCATCATCAGCGAGGCTGTGGAAATTTATATTGGAAGCTTTTGGCAAG 1277

QY 1307 AGCTAAATTTATTTCTCTTATTTACTGTAAATCATATGCTAGATCTTTTGGTTAACTGGCT 1366  
Db 1278 AGCTAAATTTATTTCTCTTATTTACTGTAAATCATATGCTAGATCTTTTGGTTAACTGGCT 1337  
QY 1367 GCATATATACCTTAATAACACAGGATTCGGGAAACAAAGGCAATTCCTGCATGTTGCTCTCCA 1426  
Db 1338 GCATATATACCTTAATAACACAGGATTCGGGAAACAAAGGCAATTCCTGCATGTTGCTCTCCA 1397  
QY 1427 TGGACCAATTTTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAAGCA 1486  
Db 1398 TGGACCAATTTTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAAGCA 1457  
QY 1487 GCTTTTGAAGCGGAAACCTGAAAGAGGTTTGAGTATCTTACAGTCTGTAATTTTGAGCG 1546  
Db 1458 GCTTTTGAAGCGGAAACCTGAAAGAGGTTTGAGTATCTTACAGTCTGTAATTTTGAGCG 1517  
QY 1547 GATAGTGATGAGCCAGCTTAATTCCTCTGAAGATTTGCCCTCAGTGGTTTAACTTTT 1606  
Db 1518 GATAGTGATGAGCCAGCTTAATTCCTCTGAAGATTTGCCCTCAGTGGTTTAACTTTT 1577  
QY 1607 TGCTGCAATCAAAATAGTACCAAGCTGCTCTCTGCTACACCATCATGTAGAGAAACA 1666  
Db 1578 TGCTGCAATCAAAATAGTACCAAGCTGCTCTCTGCTACACCATCATGTAGAGAAACA 1637  
QY 1667 TCGCCAGATGCTGCCAGTCAATTAGAGTACCGCTGGAGAGACTCAGTGCAGATCTGCAC 1726  
Db 1638 TCGCCAGATGCTGCCAGTCAATTAGAGTACCGCTGGAGAGACTCAGTGCAGATCTGCAC 1697  
QY 1727 AAAACCCGCTGACACCTTTCTCCCTTTGATCCCTGTGCTGAGAGGTCAAAGAAAT 1786  
Db 1698 AAAACCCGCTGACACCTTTCTCCCTTTGATCCCTGTGCTGAGAGGTCAAAGAAAT 1757  
QY 1787 CATGTATCTTATTTATCAGGTATGGAAGATCATGAGTCTGAGAGCTACAGGATTCAA 1846  
Db 1758 CATGTATCTTATTTATCAGGTATGGAAGATCATGAGTCTGAGAGCTACAGGATTCAA 1817  
QY 1847 GAAACCCATGAAAAAGGACATAGTGGAGATGAAGATGATGATCTTTCTGAAAGGCGAAGT 1906  
Db 1818 GAAACCCATGAAAAAGGACATAGTGGAGATGAAGATGATGATCTTTCTGAAAGGCGAAGT 1877  
QY 1907 GCCCCAGAAATGATACCGTGTATGGGATCACACCAAGCTCTTTTGACACGCAATTTCCGAAG 1966  
Db 1878 GCCCCAGAAATGATACCGTGTATGGGATCACACCAAGCTCTTTTGACACGCAATTTCCGAAG 1937  
QY 1967 TCCCTTCAAGTAGTCTGGGCTCCCGACCGCTGTTGTACATGCAACCCAGTCCCTCTGAGC 2026  
Db 1938 TCCCTTCAAGTAGTCTGGGCTCCCGACCGCTGTTGTACATGCAACCCAGTCCCTCTGAGC 1997  
QY 2027 GCAGAAATTTTGTGACTGAGATGTGACATTTGGGATTTCCCCAT 2068  
Db 1998 GCAGAAATTTTGTGACTGAGATGTGACATTTGGGATTTCCCCAT 2039

RESULT 4  
BC036182  
LOCUS BC036182  
DEFINITION Homo sapiens RNA polymerase I transcription factor RPN3, mRNA (cdna clone IMAGE:529858), partial cds.  
ACCESSION BC036182  
VERSION BC036182.1 GI:23243332  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 3709)  
AUTHORS Klausner,R.D., Collins,F.S., Wagner,L., Shennen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shennen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932  
2 (bases 1 to 3709)  
Strausberg, R.  
Direct Submission  
Submitted (31-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 63 Row: 0 Column: 6  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21361630.

## FEATURES

Location/Qualifiers  
1..3709  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5298581"  
/tissue\_type="Testis"  
/clone\_lib="NIH\_MGC\_97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"

## ORIGIN

Query Match 97.1%; Score 2008; DB 9; Length 3709;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2011; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 53 GCCTTAGTTCGGCCCAATGGCGGACCGCTGCTTCACACGCGTTGCGGGAGATGCGGC 112  
DB 2 GGGTTAGTTCGGCCCAATGGCGGACCGCTGCTTCACACGCGTTGCGGGAGATGCGGC 61  
QY 113 CGCTTCGTCTCTGCAGTTAAGAGCTGGGGGGGTCGAGGACTGGGATTTCAATATGCG 172  
DB 62 CGCTTCGTCTCTGCAGTTAAGAGCTGGGGGGGTCGAGGACTGGGATTTCAATATGCG 121  
QY 173 TGCATTAGAGAAATGACTTTTTCAATCTCCCAAGAAACCTGTCGGTTGGTGGAAAC 232  
DB 122 TGCATTAGAGAAATGACTTTTTCAATCTCCCAAGAAACCTGTCGGTTGGTGGAAAC 181  
QY 233 TGTGACAGAGTCTTGTCTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTTGTGAA 292

182 TGTGACAGAGTCTTGTCTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTTGTGAA 241  
QY 293 GAACAGAGTCTTGTAGATCCAGACATAAAGGATGACAGATCATCAACTGCTGCTGAAAT 352  
DB 242 GAACAGAGTCTTGTAGATCCAGACATAAAGGATGACAGATCATCAACTGCTGCTGAAAT 301  
QY 353 CGGTTCTTCTATCATGTACTTGACAAAAGACCTTTGAGCAACTTATCAGTATTATATAAG 412  
DB 302 CGGTTCTTCTATCATGTACTTGACAAAAGACCTTTGAGCAACTTATCAGTATTATATAAG 361  
QY 413 ATTGCTTGGTTGAATAGAAAGTCAACAGTATGGAAGAGTATTGGCTTTCTTGGTAA 472  
DB 362 ATTGCTTGGTTGAATAGAAAGTCAACAGTATGGAAGAGTATTGGCTTTCTTGGTAA 421  
QY 473 TCTTGTATCAGCACAGACGTGTTTCTCAGACCGGTCTCAGCATGATGCTTCCCATTT 532  
DB 422 TCTTGTATCAGCACAGACGTGTTTCTCAGACCGGTCTCAGCATGATGCTTCCCATTT 481  
QY 533 TGTGCTTCCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAGATTCTGATGATGA 592  
DB 482 TGTGCTTCCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAGATTCTGATGATGA 541  
QY 593 AGATGATATCTTCTCGAAAATTTTGCACATGTACAGAGCCCTTGCATAATATACGAA 652  
DB 542 AGATGATATCTTCTCGAAAATTTTGCACATGTACAGAGCCCTTGCATAATATACGAA 601  
QY 653 ATATGTACCATCGACACCGGTGTTCTCATGCCATCTGCTGGGAAAATTTCCATTTGT 712  
DB 602 ATATGTACCATCGACACCGGTGTTCTCATGCCATCTGCTGGGAAAATTTCCATTTGT 661  
QY 713 TCGAAAATCAGAGAGAACACTGGAATGTTACGTTTCACTTACTTAAGGATTAGTGATA 772  
DB 662 TCGAAAATCAGAGAGAACACTGGAATGTTACGTTTCACTTACTTAAGGATTAGTGATA 721  
QY 773 TTTTCCAACTTGGAGCATGAAATTTCTGGAGCTTATTATGAAAACCTACTCAAGTTGGA 832  
DB 722 TTTTCCAACTTGGAGCATGAAATTTCTGGAGCTTATTATGAAAACCTACTCAAGTTGGA 781  
QY 833 TGTGAATGATCCGGCAGGTTATGAAGATGCTGAAGAAAACAGCAACTCAAACTTTGTGG 892  
DB 782 TGTGAATGATCCGGCAGGTTATGAAGATGCTGAAGAAAACAGCAACTCAAACTTTGTGG 841  
QY 893 TGSGACAGATTCACCGGAAGGATTTGTTAATATGGAATGAAGATGAAGAACTGAAACATGA 952  
DB 842 TGSGACAGATTCACCGGAAGGATTTGTTAATATGGAATGAAGATGAAGAACTGAAACATGA 901  
QY 953 AACAAAGGCTGCTCGAAGCGCTCGACAGATGCTGTCCTGTCGAGCGGAGCGCTGGA 1012  
DB 902 AACAAAGGCTGCTCGAAGCGCTCGACAGATGCTGTCCTGTCGAGCGGAGCGCTGGA 961  
QY 1013 CATCTGTAGTCTTTTGGTTTGTCTTACATGAAGGATGCTGCTATGTAGATGTTAAGGT 1072  
DB 962 CATCTGTAGTCTTTTGGTTTGTCTTACATGAAGGATGCTGCTATGTAGATGTTAAGGT 1021  
QY 1073 TGTAAACGGCAAAACAAAGGATCTATATCGGACCTGATAAACAATCTTTGACAACTCCT 1132  
DB 1022 TGTAAACGGCAAAACAAAGGATCTATATCGGACCTGATAAACAATCTTTGACAACTCCT 1081  
QY 1133 GTTGCCCAACCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1192  
DB 1082 GTTGCCCAACCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1141  
QY 1193 ATTGGGATTCGACAGAGCGCATTTTGGAAACATCTCGAAAAAATTCGAGGACCAAGTAA 1252  
DB 1142 ATTGGGATTCGACAGAGCGCATTTTGGAAACATCTCGAAAAAATTCGAGGACCAAGTAA 1201  
QY 1253 TCTTGCATCATCAGCAGCGCTCTGGAATATATATGGAAGCTTTTGGCAGAGCTAA 1312  
DB 1202 TCTTGCATCATCAGCAGCGCTCTGGAATATATATGGAAGCTTTTGGCAGAGCTAA 1261  
QY 1313 ATTATTTCTCTTTTACTGTTAAATCATCCCTAGATCTTTTGTGTTAACTGGCTGCACAT 1372  
DB 1262 ATTATTTCTCTTTTACTGTTAAATCATCCCTAGATCTTTTGTGTTAACTGGCTGCACAT 1321



```
QY 1373 ATACCTTAATACCAAGGATTCCGGAAACAAAGGATTCGCGATGTGCTCTCCATGGACC 1432
Db 1382 ATACCTTAATACCAAGGATTCCGGAAACAAAGGATTCGCGATGTGCTCTCCATGGACC 1381
QY 1433 ATTTTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACAAAGCAGCTTTT 1492
Db 1382 ATTTTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGACAAAGCAGCTTTT 1441
QY 1493 GAGCGGAACCTGAAAGAGGTTTCAGTATCTTCAGAGTCTGAAATTTTGACGGGATAGT 1552
Db 1442 GAGCGGAACCTGAAAGAGGTTTCAGTATCTTCAGAGTCTGAAATTTTGACGGGATAGT 1501
QY 1553 GATGAGCCAGCTAAATCCCTGAAAGATTTCCTGCGCTCAGTGGTTAACTTTTTTGTGTC 1612
Db 1502 GATGAGCCAGCTAAATCCCTGAAAGATTTCCTGCGCTCAGTGGTTAACTTTTTTGTGTC 1561
QY 1613 AATCACAATAAGTACCAAGCTGCTCTCTGCTACACCATCATTTGAGAGGAACAATCGCCA 1672
Db 1562 AATCACAATAAGTACCAAGCTGCTCTCTGCTACACCATCATTTGAGAGGAACAATCGCCA 1621
QY 1673 GATGCTGCAGTCAATTAGAGATGACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACC 1732
Db 1622 GATGCTGCAGTCAATTAGAGATGACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACC 1681
QY 1733 GCTGGACACCTTCTCCCTTTGATCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1792
Db 1682 GCTGGACACCTTCTCCCTTTGATCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1741
QY 1793 TCCTATTTATCAGGTATCGGGAAGACATCAGTGTCTGAAGAGCTACAGGAGTTTCAGAAACC 1852
Db 1742 TCCTATTTATCAGGTGTGGGAAGACATCAGTGTCTGAAGAGCTACAGGAGTTTCAGAAACC 1801
QY 1853 CATGAAAAGACATPAGTGGGAAGATGAAGATGATGACTTTCTGAAAGCGGAAGTGCCTCCA 1912
Db 1802 CATGAAAAGACATPAGTGGGAAGATGAAGATGATGACTTTCTGAAAGCGGAAGTGCCTCCA 1861
QY 1913 GAATGATACCTGTGTTGGATCACCAAGCTCCTTTGACAGCGATTTCCGAAGTCCTTC 1972
Db 1862 GAATGATACCTGTGTTGGATCACCAAGCTCCTTTGACAGCGATTTCCGAAGTCCTTC 1921
QY 1973 AAGTAGTGTGGCTCCCAACCGCTGTTGTACATGCAACCCAGTCCCTCTACGCGCAGAA 2032
Db 1922 AAGTAGTGTGGCTCCCAACCGCTGTTGTACATGCAACCCAGTCCCTCTACGCGCAGAA 1981
QY 2033 ATTTGTGACTGAGATGACATTTGGGATTCCTCCAT 2068
Db 1982 ATTTGTGACTGAGATGACATTTGGGATTCCTCCAT 2017

RESULT 5
AC012547
LOCUS Homo sapiens clone RP11-11K1, WORKING DRAFT SEQUENCE, 14 unordered
DEFINITION AC012547 135873 bp DNA linear HTG 30-MAR-2000
pieces.
ACCESSION AC012547
VERSION AC012547.2 GI:7341888
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 135873)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouthgatter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
```

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehoczky,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McQuirk,A., McKernan,K., McLaughlin,J., Meldrum,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,I., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL

## COMMENT

Submitted (29-OCT-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 30, 2000 this sequence version replaced gi:6139119.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/JM/RepeatMasker.html  
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3178

Center clone name: ll\_K1

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 127596 bases at least Q40

Consensus quality: 130929 bases at least Q30

Consensus quality: 132388 bases at least Q20

Insert size: 133000; agarose-fp

Insert size: 134573; sum-of-contigs

Quality coverage: 4.9 in Q20 bases; agarose-fp

Quality coverage: 4.9 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1516: contig of 1516 bp in length  
1517 1616: gap of 100 bp  
1617 3136: contig of 1520 bp in length  
3137 3236: gap of 100 bp  
3237 5148: contig of 1912 bp in length  
5149 5248: gap of 100 bp  
5249 7742: contig of 2494 bp in length  
7743 7842: gap of 100 bp  
7843 10650: contig of 2808 bp in length  
10651 10750: gap of 100 bp  
10751 16194: contig of 5444 bp in length  
16195 16294: gap of 100 bp  
16295 25751: contig of 9457 bp in length  
25752 25851: gap of 100 bp  
25852 37209: contig of 11358 bp in length  
37210 37309: gap of 100 bp  
37310 47828: contig of 10519 bp in length  
47829 47928: gap of 100 bp  
47929 64239: contig of 16311 bp in length  
64240 64339: gap of 100 bp  
64340 77846: contig of 13507 bp in length  
77847 77947: gap of 100 bp  
77948 95311: contig of 17365 bp in length  
95312 95411: gap of 100 bp  
95412 108930: contig of 13519 bp in length  
108931 109030: gap of 100 bp  
109031 135873: contig of 26843 bp in length.

Location/Qualifiers

1. 135873

/organism="Homo sapiens"

FEATURES  
source



```
Db      98269 TTGCTGCTGCACGAATAATATACAGTAG---TCTGCTACACTATCATTTGAGAGGAACA 98325
QY      1666 ATGCGCAGATGTCGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCA 1725
Db      98326 ACGGCAGAG-----AATACAGCTGGAGGAGACTCGGTGTGAACCTGCA 98368
QY      1726 CAACCCGCTGACACCTTCTCCCTTTGATCCCTGTGTGCTGAAGAGCTCAAGAAT 1785
Db      98369 CCAACCCATTGACGCGCTTCTCCCTTTGATCCCTGTGTGCTGAAGAGCTCAAGAAT 18428
QY      1786 TCATTGATCTATTATATCAGGTATGGGAAGACATGAGTCTGAAGAGCTACAGGAGTTCA 1845
Db      98429 TCATTGATCTATTATATCAGGTATGGGAAGACATGAGTCTGAAGAGCTACAGGAGTTCA 98488
QY      1846 AGAACCCTATGAAAAGGACATAGTGGAGATGAAGATGATGATCTTCTGAAGGCGAAG 1905
Db      98489 AGAACCCTATGAAAAGGAGATGAAAAGATGGAGATGAT---CTTTTGAAGGTGAAG 98545
QY      1906 TGCCCCAGATGATACCGTGATGGGATCACACCAAGCTCCCTTGCACAG-----CATTT 1960
Db      98546 TTCCCCCAATGATACCGATGCGGATTTATACCGAGCCACTTTGATGCGATGCACATTT 98605
QY      1961 CCGAGTCTCTCAAGTAGTGTGGGCTCCGCCACCCCTGTTGTACATG-CAAACCCAGTCCCC 2019
Db      98606 CCAAGTCTCTCAAGTAGTGTGGGCTCCGCCACCCCTGTTGTACATG-CCAAGACCACTTCTT 98665
QY      2020 TCTGAGGCGCAAAATTTGTGACTGATGATGTGACATT 2055
Db      98666 TCTCATCTCGAGGATTTGTGATTGATGATGCAGCAT 98701

RESULT 6
AC013444      192815 bp      DNA      linear      PRI 09-JAN-2002
DEFINITION   Homo sapiens BAC clone RP11-470B22 from 2, complete sequence.
AC013444
VERSION      AC013444.8 GI:18093267
KEYWORDS     HTG.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 192815)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED      9847074
REFERENCE    2 (bases 1 to 192815)
Waligorski,J., Haakenson,W. and Spalding,L.
The sequence of Homo sapiens BAC clone RP11-470B22
Unpublished (2001)
JOURNAL     3 (bases 1 to 192815)
Waterston,R.H.
Direct Submission
Submitted (11-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 192815)
Waterston,R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 9, 2002 this sequence version replaced gi:13518264.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
-----
Summary Statistics
Center project name: H_NH0470B22
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenio,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC024616; the clone sequenced to the right is RP11-190J23. Actual start of this clone is at base position 1 of RP11-470B22; actual end is at base position 192815 of RP11-470B22.

Unresolved homopolymeric run from base position 164727 to 164753.

FEATURES	source
repeat_region	1. 192815
repeat_region	/organism="Homo sapiens"
repeat_region	/mol_type="genomic DNA"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="2"
repeat_region	/map="2"
repeat_region	/clone="RP11-470B22"
repeat_region	/clone_lib="RPCI-11"
repeat_region	1. 575
repeat_region	/rpt_family="Li"
repeat_region	611. 878
repeat_region	/rpt_family="ERV1"
repeat_region	879. 1032
repeat_region	/rpt_family="(TAGA)n"
repeat_region	1046. 1092
repeat_region	/rpt_family="ERV1"
repeat_region	1987. 2346
repeat_region	/rpt_family="L2"
repeat_region	2737. 2972
repeat_region	/rpt_family="L1"
repeat_region	3702. 3739
repeat_region	/rpt_family="AT-rich"
repeat_region	4870. 4891
repeat_region	/notes="match to EST BE071256 (NID:98416143)"
repeat_region	6880. 6992
repeat_region	/rpt_family="MIR"
repeat_region	8677. 8934
repeat_region	/rpt_family="Alu"
repeat_region	8773. 8800
repeat_region	/rpt_family="AT-rich"
repeat_region	8955. 8989
repeat_region	/rpt_family="AT-rich"
repeat_region	9207. 9247
repeat_region	/rpt_family="L2"
repeat_region	9342. 9773
repeat_region	/rpt_family="L2"

repeat_region	9977..10011	/rpt_family="TTTTA)n"	20592..20963	misc_feature	/note="similar to Homo sapiens EST AA481295 (NID:g2210847)
repeat_region	10976..11265	/rpt_family="Alu"	20703..21362	misc_feature	/note="similar to Homo sapiens EST BG502871 (NID:g13464388)"
repeat_region	12178..12207	/rpt_family="(T)n"	20798..21056	misc_feature	/note="similar to Homo sapiens EST AV649508 (NID:g9870522)"
repeat_region	12362..12390	/rpt_family="AT_rich"	20862..21269	misc_feature	/note="similar to Homo sapiens EST AA191111 (NID:g1779805)
repeat_region	12702..12731	/rpt_family="(TG)n"	20930..21593	misc_feature	/note="similar to Homo sapiens EST BF203684
repeat_region	12731..12758	/rpt_family="(GA)n"			
repeat_region	12758..12785	/rpt_family="(TG)n"			
repeat_region	13033..13054	/rpt_family="(CAAAA)n"			
repeat_region	13195..13699	/rpt_family="L1"			
repeat_region	13543..13567	/rpt_family="AT_rich"			
repeat_region	14142..14305	/rpt_family="MIR"			
repeat_region	14500..14534	/rpt_family="AT_rich"			
repeat_region	14617..14735	/rpt_family="MIR"			
repeat_region	14736..14836	/rpt_family="MER2_type"			
repeat_region	14837..15106	/rpt_family="Alu"			
repeat_region	15107..15317	/rpt_family="MER2_type"			
repeat_region	15318..15385	/rpt_family="MIR"			
repeat_region	15578..16066	/rpt_family="ERV1"			
repeat_region	16072..16112	/rpt_family="AT_rich"			
repeat_region	16192..18222	/rpt_family="AT_rich"			
repeat_region	18612..18899	/rpt_family="Alu"			
misc_feature	19181..19485	/note="match to EST R89091 (NID:g953918) yq02f04.s1"			
misc_feature	19879..21714	/note="similar to Rattus norvegicus EST Y16973 (NID:g4727035)"			
misc_feature	19890..20284	/note="match to EST R89197 (NID:g954024) yq02f04.r1"			
misc_feature	19915..20369	/note="similar to Homo sapiens EST BG535886 (NID:g13527431)"			
misc_feature	19923..20378	/note="similar to Homo sapiens EST AA213789 (NID:g1812416) zr91a11.s1"			
misc_feature	19923..20369	/note="similar to Homo sapiens EST BG428305 (NID:g13334811)"			
misc_feature	20091..20378	/note="similar to Homo sapiens EST AI742153 (NID:g5110441) wg49f01.x1"			
misc_feature	20158..20664	/note="similar to Homo sapiens EST AV703279 (NID:g10720608)"			
misc_feature	20243..21062	/note="similar to Mus musculus EST BG295610 (NID:g13057417)"			
misc_feature	20412..21044	/note="similar to Homo sapiens EST BE734354 (NID:g10148346)"			
misc_feature	20567..20666	/note="similar to Homo sapiens EST BG778634 (NID:g14048951)"			
<hr/>					
Query Match	58.6%	Score 1211.6	DB 9	Length 192815	
Best Local Similarity	80.2%	Pred. No. 0	Mismatches 334	Indels 65	Gaps 14
Matches 1617	Conservative	0			
QY	60	TTTGGCCCAATGGCGGACCGCTGTTTCACACGCGTTTCCCGGAGATGCGCGCTTCG	119		
Db	19803	TTAAGCCCGGGCAACAGGCTGCTTCTCACATGTTGCGGG- GAAGCAGAGGCTTCA	19861		
QY	120	TCCTCTGCAGTTPAAGAAGCTGGCGCTGCGAGACTGGGATTTCAAATATGCGTCAATTA	179		
Db	19862	GCTTCTGTGTCAGAAGCTGGGTATGTCATGCACAGTATTTTCAGATATGCTTGCAATTC	19921		
QY	180	GAGAATGACTTTTCAATTCCTCCCAAGAAAACCTGTCGGTTTGGTGAAGCTGACA	239		
Db	19922	AAAGAATGATTTTTCAGCTCTCCCAAGAAAATTTGTCATCTGTGATGAAACGGGGACA	19981		
QY	240	GAAGTCTTGTCTGAAGTACAAAAGGCTGAAACAAATGACTTTGAGTTGTTGAAGAACCCAG	299		
Db	19982	GAAGTCTTGTCCGAGGACAAAGAGTGAACAAATTTAACTTTGAGCTGTTGAAGACCAG	20041		
QY	300	CTGTTAGATCCAGACATAAAGGATGACAGATCATCAACTGGCTGCTAGAAATTCGGTTCT	359		
Db	20042	TTGTGAGATTCAAACATAAAGGATGACAGATCTTGAACCTGGCTGTGAAAATTTTGTCT	20101		
QY	360	TCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATATATATTAAGATGCGCT	419		
Db	20102	TCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATATATATTAAGATGCGCT	20161		
QY	420	TGTTTGAATAGAGTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCTTGGTAACTTTGTA	479		
Db	20162	TGTTTGAATAGAGTCAAAACAGTAGTGAAGAGTATTTGGCTTTTCTTGGTAACTTTGTA	20221		
QY	480	TCAGCAGACACTGTTTTCTCTCAGACCGTGTCTCAGCATGATTCGTTCCCAATTTTGTG---	536		
Db	20222	ACAACAGA---CTCTTCTCTAGCCCAAGTCTCAGCATAAATTTGCTTCCCAATTTGGCGTAC	20277		
QY	537	CCTCCCGGAGTGTATCAATTAAGNAAGCGGATGTAGATGTTTCAGATTCGTATGATGAAGAT	596		
Db	20278	CCTCCCTGGAATGATCAATTAAGNAAGCGGATGTAGATGTTTCAGATTCGTATGATGAAGAT	20337		
QY	597	GATAATCTTCTCCGAAATTTTGACACATGTCAAGAGCCCTTGCATAATATAGCAAGATAT	656		
Db	20338	GATTATCTTCTGACATTTTGACACAGTCAACCTTACAAATTAACAGCAAGATAT	20397		
QY	657	GTACCATGACACCGGTGTTTCTCATGCCAATCTCGTGGGAAAAAATTCATTTGTTGCGA	716		
Db	20398	GTCCCACTGGCACTGGTTTCTTTTGGCCCATCTCGTGGGAAAAAATTCATTTGTTGCGA	20455		
QY	717	AAATCAGAGAGAACCTGGAATGTTTACGTTTCAATCTACTAAGGATTAGTGTATATTTT	776		
Db	20456	-AATCAGAGAGAACCTGGAATGTTTACATTCATACTTACTAAGATTAGTATATATCTTT	20514		
QY	777	CCAACTTGGAGGATGAAATTTCTGGAGCTTTATTTGAAAAAACTACTCAAGTTGGATGTG	836		
Db	20515	CCAACTTGGAGGATGAAATTTCTGAAGCTTTATTTAAATAATCTATTCAAAATGCGATGTG	20574		
QY	837	AATGCAATCCCGGAGGGTATTTGAAGATGCTGAAGAAACACACAACTCAAACTTGCGTGGG	896		

Db 20575 AATACATTCTGGCATGATATTGAAGTCTGAGGAAACAGCGCTCAAACTTGTGTGGTGG 20634  
QY 897 ACAGATTCCAGGAGAGATTGTTTAAATGCGATGAAGATGAAGAACTGAACATGAAACA 956  
Db 20635 ACAGATTCCAGGATGATGTTTCAATATGAACGAAGATGAAGAGAC-----TGTTTCT 20688  
QY 957 AAGGCTGGTCCCTGAACGGCTCCAGCAGATGCTGATCTGATGAGCGCGCTGACATC 1016  
Db 20689 AAGGCTGACTCAAAATGGCTCGGTGAGATGGTGCACCCCTGTAGCTGAATGCGCTGACACC 20748  
QY 1017 CTGATGCTTTGGTTTGTCTACATGAAGGATGCTGCTATG-----TAGATGTTAAGG 1071  
Db 20749 CTGCTGCTTTTGTCTTTGTCAGTACTTGAAGTATGCTCTATGTAATAATCAATGTAAGG 20808  
QY 1072 TTGATAACGGCAAAACAAAGAGTATATATGCGACCTGATAAACAATCTTTTGCAAACTCC 1131  
Db 20809 TTGATAACAAACAAATAAAGGATTTATA-----ATATATCACAAACTGATAACATA 20860  
QY 1132 TGTTCGCCACCATGCTCCTGCCATGTACAGTCTTTTTCATGTTTACCTCTGTAGTTTCA 1191  
Db 20861 TTTGCAAAACCCCTGCCCTCTGTCAGTACAGTCTTTTATGTTTATGTTTACCTCTGTGCTTAA 20920  
QY 1192 AATTGGGATTCGACGAGCATTTTGGAAACATCTCTGG---AAAAAATTGCAGGACCCAA 1248  
Db 20921 AATTGCGACTCTCAGAAGCATTTTGGAACTCTCTGGAAAAAATAATGAGGATCCAA 20980  
QY 1249 GTAATCCTGCCATCATACAGGAGGCTGCTGGAATATATATGGAAGCTTTTGGCAAGAG 1308  
Db 20981 ATAACTCCTGTGATCATCATCAGGTGAGTCTGCGAAATATATATGGAAGCTCTTTGGCAAAAG 21040  
QY 1309 CTAATATTATTCCTCTTATTT---ACTGTAAATCATGCTGAGATCTTTTGGTAACTGGC 1365  
Db 21041 CTAATATTATTCCTCTCATTAACAATATAAATCATGCTAGGCTTTTCGGTTAACTGGC 21100  
QY 1366 TGCATATATACCTTAATAACAGGATTCGGAAACAAAGGCATTCGCGATCTGTGCTCC 1425  
Db 21101 CACAAATACCTTACTTAACACAGATCTGAGAAATGAGGCTTTTGTGATATGCTCTCC 21160  
QY 1426 ATGACACATTTACTCAGCTGCGAAGCTGTGTTCTACACCTTTTGTATAGACACAAGC 1485  
Db 21161 ATCAAAATTTTACTCAGCTTCCAAAGCTGTGTTCCACACCTTTTGTATAGACACAAGC 21220  
QY 1486 AGCTTTTGGCGGAAACCTGAAGAGTTTGCAGTATCTTCAGATCTGATTTTGAAGC 1545  
Db 21221 AACTTTTAAAGCTAAACCTGAAGAGTTTGAAGTGTCTTCAGTGCCTACATTTTAAAG 21280  
QY 1546 GGATAGTATGAGCCAGCTAAATCCCTGAAGATTTGCTGCCCTCAGTGGTTAACTTTT 1605  
Db 21281 GGATAGTATGAGCCAGCTAAATCCCTGAATATTTGCTGCCCTCAGTGGTTAACTTTAT 21340  
QY 1606 TTGCTGCAATCAAAATAAGTACAGCTGCTCTTCTGCTACACATCATTTGAGAGGAACA 1665  
Db 21341 TTGCTGCTGTCAAGATAAATATCAGCTAG---TCTGCTACACTATCATTTGAGAGGAACA 21397  
QY 1666 ATGCCAGATGCTGCCAGTCAATAGGAGTACCGCTGGAGAGACTCAGTGCAGATCTGCA 1725  
Db 21398 ACGCCAGA-----AATACAGCTGGAGGAGACTCGGTGTGAACGTGCA 21440  
QY 1726 CAACCCGCTGGACACCTTCTCCCTTTTGCATCCCTGTGCTGCTGAAGAGTCAAGAAAT 1785  
Db 21441 CCAACCCATGGAGCGCTTCTCCCTTTTGCATCTTGTGCGCTTAAAGTCAAGTCAAGAAAG 21500  
QY 1786 TCATTGATCTATTATCAGGTATGGGAAGACATGAGTGTGTAAGACTACAGAGTTCA 1845  
Db 21501 TCATTGATCTATTATCAGATATGGGAAGACATAAGTGTGGAAGAGCTTTCAGGAGTTTA 21560  
QY 1846 AGAAACCATGAAAAAGGACATATGGAAGATGAAGATGATGATTTCTGAAAGGCGAAG 1905  
Db 21561 AGAAACCTATGAAAAAGGAGATGAAAAAGATGGAGATGAT---CTTTTGAAGGTGAAG 21617  
QY 1906 TGCCCCAGAAATGATACCGTATTTGGGATCACCAAGCTCTTTTGAACAG-----CATTT 1960  
Db 21618 TTCCCCAAATGATACCATGATCGGATTTATACGAGCCACTTTTGTGATGCGATGCACATTT 21677

QY 1961 CGAAGTCTCTCAAGTAGTGTGGCTCCCGCTCCCGCTGTTGTACATG-CAACCCAGTCCCC 2019  
Db 21678 CCAAGTCTCTCAAGTAGTGTGGCTCCCGCTCCCGCTGTTGTACACCTGCCAGACAGTCTCT 21737  
QY 2020 TGTGACGCGCAAAATTTTGTGACTGAGATGTGCATTT 2055  
Db 21738 TCTCATCTCGAGGATTTGTGATTGAGATGCACATTT 21773

RESULT 7  
BC034110  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)  
Mus musculus  
Mus musculus expressed sequence AL023001, mRNA (cDNA clone IMAGE:4504751), partial cds.  
BC034110  
BC034110.1 GI:21706612

REFERENCE  
AUTHORS  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H.K., Moore, T., Max, S.I., Wang, J., Heide, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, K.J., Mullahy, S.J., Bosak, S.A., McEwan, P.J., Richardson, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, D.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Huetten, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932  
2 (bases 1 to 3166)  
Strausberg, R.  
Direct Submission  
Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapsh-r@mail.nih.gov](mailto:cgapsh-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxli.stanford.edu](mailto:mcd@paxli.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

REMARK  
COMMENT  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 44 Row: 9 Column: 5  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

Location/Qualifiers

FEATURES

	source	1.. 3166 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:4504751" /tissue type="Eye, retina, mouse strain C57Bl\6" /clone_lib="NIH MGC_94" /lab_host="DH10B" /note="Vector: pCMV-SPORT6" cl. .3166 /gene="AL023001" /db_xref="LocusID:106298" /db_xref="MGI:2146525" <1.. .1544 /gene="AL023001" /product="AL023001 protein" protein_id="AAH34110.1" /db_xref="GI:21706613" /db_xref="LocusID:106298" /db_xref="MGI:2146525" /translation="PCLSMASHFVPVRVIVKEGDIVSDDEDDMLPALFDTCHRA LOIITRYVPSWFLFMPIEDIAEKFPVKRSRTLECYVHNLRISLVPTTTRRILBY IEKLKLDVSQRDIDEAEKAADTCGGTTTEGLFNMEDEDTPEKKADQEQNP MAHPAERLDVLICLLSVIEDVCRVHGKI DNNTKOLYKDLISI PKLLLPHTASCH VQFFPFELCSFKLGAEAPLEHLWKKLQDPNNPALIRQAANYIGSFARAKIPILIT VKTCDLLVNLWHMLYTDSKTKAGLKQDPNPNALIRQAANYIGSFARAKIPILIT KGQLQYLQSINFETVLSQANPKICLPQVNVFAAITNKYQLVCYTIMERSRMRL PVIRSPTAGDSVQTCTNPILDTPFDPCVLKRKKFI DPIQIWEDGSABELEGFEKKS TKVEVDEDDDDFLNGEVPSQSDTVTLTPSPDFTHFSQSPSSSVSGPPVLVI PQSQPLL TRYLD"
	ORIGIN	
	Query Match	57.1%; Score 1179.8; DB 10; Length 3166;
	Best Local Similarity	85.3%; Pred. No. 0;
	Matches 1327; Conservative	0; Mismatches 227; Indels 1; Gaps 1;
2y	502	GACGGTGTCACATGATTGCTCCCATTTTTGTGCTCCCGAGTGATCAATAAGAAG 561
b	1	GACCATGTCTCACATGATTGGGTCTCATTTTTGTACTCTCCCGAGTAATTGTCAAGGAAG 60
y	562	GCATGTAGTAGTGTTCAGATTCTGATGATGAAGATGATAATCTTCCTCAAATTTTGACA 621
b	61	GTGCCATAGATGTTTCAGATTCTGATGACGAAGATGACAACTTCTCCTCAATTTTGACA 120
y	622	CATGTCTACAGAGCCTTGCAATAATAGCAAGATATGTACCATFCGACACGGTGGTTTCTCA 681
b	121	CATGTCTACAGAGCCTTGCAATAATAAACAAAGATATGTCTCCATFCGACACATGGTTTCTAA 180
y	682	TGCCAATFACTGGTGGAAAAATTTCCATTTGTTTCGAAAATCAGAGAGACACTGGAACTGT 741
b	181	TGCCAATFACTGGTGGAAAAATTTCCCGTTTGTGAGGAAGTCCGAGAGAACATTGGAAATGTT 240
y	742	ACGTTTCATACTTACTAAGGATTAGTGATATATTTTCCAACCTTGAAGGCATGAATTTCTGG 801
b	241	ATGTTTCATACCTTATTAAGATTAAGTTTATATTTCCCACTTTGAGCGGTGAATTTCTGG 300
y	802	AGCTTATTTATGAAAACTACTCAAGTGTGATGTGAATGCATCCCGCAGGGTATTTGAAG 861
b	301	AGCTTGTATTGAAAAAGCTACTCAAGTGTAGATGTGAGTGTATCTGCGCGCAGGATATTGAAG 360
y	862	ATGCTGGAAGAACAGCACTCAAACTTGGTGGGACAGATTCCACGGAAGGATTGTTTAT 921
b	361	ATGCTGGAAGAAAGCAGCTTCACACTTGGTGGGACAGATTACCAAGAGCACTGTTTAT 420
y	922	ATATGCAATGAAGATGAAGAACTGAAATCAATCAAAAGGCTGGTCTCTGAACCGCTCGACC 981
b	421	ATATGCAATGAAGATGAAGAACACTGACCTGAGAAAGAGCTGACACGAGCAGCCTAAC 480
y	982	AGATGGTGCACTCTGTATAGCCGAGCGCTTGACATACCTCTGATGTCTTTGGTTTGTCTCTACA 1041
b	481	AGATGGCTCATCCACACCGCAGAGCGCTGAGACCTCTGCTGTCTTCTGCTGTCTCTACA 540

```
ACCESSION AX427479
VERSION AX427479.1 GI:21537625
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Hodgson,D.M., Lincoln,S.E., Russo,F.D., Spiro,P.A., Banville,S.C.,
Bratcher,Shawn,R., Dufour,G.F., Cohen,H.J., Rosen,B.H., Shah,P.,
Chalup,M.S., Hillman,J.L., Jones,A.L., Yu,J.Y., Greenawalt,L.B.,
Panzer,S.R., Roseberry,A.M., Wright,R.J., Chen,W., Liu,T.,
Yap,P.R., Stockdreher,T.K., Ameshey,S. and Fong,W.T.
HUMAN secretory molecules
TITLE Human secretory molecules
JOURNAL Patent: WO 0123558-A 52 05-APR-2001;
Incyte Genomics, Inc. (US)
FEATURES
    source
        1..1423
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            /note="Incyte ID No: 428745.2.dec"
ORIGIN
    Query Match 43.2%; Score 894; DB 6; Length 1423;
    Best Local Similarity 80.9%; Pred.No.2.6e-235;
    Matches 1191; Conservative 0; Mismatches 10; Indels 271; Gaps 2;
QY 157 GGATTTCAAATATGGTGCATTAGAGATGACTTTTTCAAATCTCTCCCAAGAAACCTG 216
DB 192 GGAATTTCAAATATGGTGCATTAGAGATGATTTTTCAAATCTCTCCCAAGAAACCTG 251
QY 217 TTCGGTTTGGTGGAACTGTGACAGAGTCTTCTGGAAGTCAAAAAGGGTGAACAAATG 276
DB 252 TTCGGTTTGGTGGAACTGTGACAGAGTCTTCTGGAAGTCAAAAAGGGTGAACAAATG 311
QY 277 ACTTTGAGTCTGTGAAGAACAGCTGTAGATCCAGACATAAAGATGACAGATCATCA 336
DB 312 ACTTTGAGTCTGTGAAGAACAGCTGTAGATCCAGACATAAAGATGACAGATCATCA 371
QY 337 ACTGGCTGTAGAAATTCGGTTCTTCTATCATGTACTTGACAAAAGACTTTGACAACTTA 396
DB 372 ACTGGCTGTAGAAATTCGGTTCTTCTATCATGTACTTGACAAAAGACTTTGACAACTTA 431
QY 397 TCAGTATTATTAAGATGTCCTTGGTTGAATAGAGTCAACAGTAGTGAAGAGTATT 456
DB 432 TCAGTATTATTAAGATGTCCTTGGTTGAATAGAGTCAACAGTAGTGAAGAGTATT 491
QY 457 TGGCTTTCTTGGTAATCTGTATCAGCAGACACTGTTTCTCAGACCGTGTCTCAGCA 516
DB 492 TGGCTTTCTTGGTAATCTGTATCAGCAGACACTGTTTCTCAGACCGTGTCTCAGCA 551
QY 517 TGAATGCTTCCATTTTGTGCTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTT 576
DB 552 TGAATGCTTCCATTTTGTGCTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTT 611
QY 577 CAGATTCATGATGAAGATGATTAATCTTCTGCAAAATTTTGACACATGTCACAGAGCCT 636
DB 612 CAGATTCATGATGAAGATGATTAATCTTCTGCAAAATTTTGACACATGTCACAGAGCCT 671
QY 637 TGCATAATAGCAAGATATGTACCATCGACACCGTGTCTCATGCCAATATCTGTGG 696
DB 672 TGCATAATAGCAAGATATGTACCATCGACACCGTGTCTCATGCCAATATCTGTGG 731
QY 697 AAAAATTTCCATTTGTTTCGAAAAATCAGAGAACTGGAATGTTAGTTCATAAATCTAC 756
DB 732 AAAAATTTCCATTTGTTTCGAAAAATCAGAGAACTGGAATGTTAGTTCATAAATCTAC 791
QY 757 TAAGGATAGTGATATATTTTCCAACTTGAGGCAATGAAATTTCTGGAGCTTATTATTGAAA 816
DB 792 TAAGGATAGTGATATATTTTCCAACTTGAGGCAATGAAATTTCTGGAGCTTATTATTGAAA 851
QY 817 AACTACTCAAGTTGGATGTGAATGCATCCCGGCGAGGGTATTGAAGATGCTGAAGAAACAG 876
```

```
Db 852 AACTACTCAAGTTGGATGTGAATGCATCCCGCAGGGTATTGAAGATGCTGAAGAAACAG 911
QY 877 CAACCTCAAACTTGTGTGGACAGATTCCACGGAAGGATTGTTTAATATGGAATG 936
Db 912 CAATCAAACTTGTGTGGACAGATTCCACGGAAGGATTGTTTAATA 959
QY 937 AAGMAACTGAAATGAAACAAAGGCTGTCTGTAACGGCTCGACCAGATGGTGCATCCTG 996
Db 960 959
QY 997 TAGCGAGCGCTGGACATCCTGATGTCCTTGGTTTGTCTTACATGAAGGATGTCGTCT 1056
Db 960 959
QY 1057 ATGTAGATGGTAAGTTGATTAACGGCAAAACAAGGATCTATATCGCACCTGATAACA 1116
Db 960 959
QY 1117 TCTTTGCAAACTCCTGTTGCCACCCATGCTCCTGCCATGTACAGTCTTTTCATGTTT 1176
Db 960 959
QY 1177 ACCTCTGTAGTTTCAAAATGGGATTCGCAGAGG-CATTTTGGAAATCTCTCGAAAAA 1235
Db 960 1001
QY 1236 TTGCAGGACCAAGTAATCTCTGATCATCAGGAGGCTGCTGGAATATATATTGGAAGC 1295
Db 1002 TTGCAGGATCCAAATCTCTGATCATCAGGAGGCTGCTGGAATATATATTGGAAGC 1061
QY 1296 TTTTGGCAAGAGCTAAATTTATTTCTTAAATCTGTAATATCATGCTAGATCTTTTGG 1355
Db 1062 TTTTGGCAAGAGCTAAATTTATTTCTTAAATCTGTAATATCATGCTAGATCTTTTGG 1121
QY 1356 GTTAACTGGCTGCACATATACCTTTAATACAGGATTCGGGAACAAAGGCAATCTTGCAT 1415
Db 1122 GTTAACTGGCTGCACATATACCTTTAATACAGGATTCGGGAACAAAGGCAATCTTGCAT 1181
QY 1416 GTTGCCTTCATGACCAATTTTACTGAGCTGCCAGTGTGTTCTACACCTTTGTTTTT 1475
Db 1182 GTTGCCTTCATGACCAATTTTACTGAGCTGCCAGTGTGTTCTACACCTTTGTTTTT 1241
QY 1476 AGACACAAGCAGCTTTTGGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGCTG 1535
Db 1242 AGACACAAGCAGCTTTTGGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGCTG 1301
QY 1536 AATTTGAGCGGATGATGATGAGCCAGCTAAATCCCTGAAAGATTTGCTGCCCTCAGTG 1595
Db 1302 AATTTGAGCGGATGATGATGAGCCAGCTAAATCCCTGAAAGATTTGCTGCCCTCAGTG 1361
QY 1596 GTTAACTTTTGTGTCGAATCACAATAAGTA 1627
Db 1362 GTTAACTTTTGTGTCGAATCACAATAAGTA 1393
```

```
RESULT 9
AX713860 LOCUS AX713860
DEFINITION Sequence 544 from Patent EPI293569.
ACCESSION AX713860
VERSION AX713860.1 GI:29888786
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahara,K. and
Masuko,Y.
Full-length cDNAs
```



JOURNAL Patent: EP 1293569-A 544 19-MAR-2003;  
Helix Research Institute (JP) ; Research Association for  
Biotechnology (JP)  
FEATURES  
source Location/Qualifiers  
1. .2493  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 27.4%; Score 567; DB 6; Length 2493;  
Best Local Similarity 98.3%; Pred. No. 6.6e-145;  
Matches 573; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1196 GGGATTCGAGAGGATTTTGGACATCTCTGGAAAAATTCGAGGACCCAGTAATCC 1255  
DB 711 GGGATTCGAGAGGATTTTGGACATCTCTGGAAAAATTCGAGGATCCAGTAATCC 770  
QY 1256 TGCCATCATCAGGAGCTGTGGAAATATATATTCGAAGCTTTTGGCAAGAGCTAAAT 1315  
DB 771 TGCCATCATCAGGAGCTGTGGAAATATATATTCGAAGCTTTTGGCAAGAGCTAAAT 830  
QY 1316 TATTCCTCTTATCTGTAATAATCATGCTAGATCTTTGGTTAACTGGCTGCACATATA 1375  
DB 831 TATTCCTCTTATCTGTAATAATCATGCTAGATCTTTGGTTAACTGGCTGCACATATA 890  
QY 1376 CTTTATTAACAGGATTCGGAAACAAAGGCATCTGCGATGTGCTCTCCATGGACCAT 1435  
DB 891 CTTTATTAACAGGATTCGGAAACAAAGGCATCTGCAATGTGCTCTCCATGGACCAT 950  
QY 1436 TTACTCAGCTCCGAGCTGTGTCTTACACCTTTGTTTATAGACAAACAGCTTTTGAG 1495  
DB 951 TTACTCAGCTCCGAGCTGTGTCTTACACCTTTGTTTATAGACAAACAGCTTTTGAG 1010  
QY 1496 CGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTCAATTTTGCAGGATAGTAT 1555  
DB 1011 CGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTCAATTTTGCAGGATAGTAT 1070  
QY 1556 GAGCGAGCTTAATCCCTGAAGATTTGCTGCTCCCTCAGTGGTTAACTTTTGGCTGCAAT 1615  
DB 1071 GAGCGAGCTTAATCCCTGAAGATTTGCTGCTCCCTCAGTGGTTAACTTTTGGCTGCAAT 1130  
QY 1616 CACAATAAGTACCACTGCTCTTCTGCTACACCATCATTCAGAGGAAACATCGCCAGAT 1675  
DB 1131 CACAATAAGTACCACTGCTCTTCTGCTACACCATCATTCAGAGGAAACATCGCCAGAT 1190  
QY 1676 GTGCGAGCTATTAGAGTACCGCTGGAGAGACTCAGTGCAGATCTGCACAAACCCGCT 1735  
DB 1191 GTGCGAGCTATTAGAGTACCACTGGAGGAGACTCAGTGCAGATCTGCACAAACCCACT 1250  
QY 1736 GGNACACTTCTCCCTTTGATCCCTGTGCTGCTGAAGAGGTCA 1778  
DB 1251 GGACACTTTCGTCCCTTTGATCCCTGTGCTGCTGAAGAGGTGA 1293  
RESULT 10  
AK055742  
LOCUS 2493 bp mRNA linear PRI 01-AUG-2002  
DEFINITION Homo sapiens cDNA FLJ31180 fis, clone KIDNE2000266.  
ACCESSION AK055742  
VERSION AK055742.1 GI:16550546  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1  
AUTHORS Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,  
Ono,Y., Horita,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,  
Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,  
Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,  
Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,  
Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,

Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,  
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,  
Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2493)  
Isogai,T., Otsuki,T. and Sugiyama,T.  
Direct Submission  
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu Chiba 292-0812, Japan  
(E-mail:genomicshri.co.jp, tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.  
Location/Qualifiers  
1. .2493  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="KIDNE2000266"  
/tissue\_type="kidney"  
/clone\_lib="KIDNE2"  
/note="Cloning vector: pME18SFL3"  
ORIGIN  
Query Match 27.4%; Score 567; DB 9; Length 2493;  
Best Local Similarity 98.3%; Pred. No. 6.6e-145;  
Matches 573; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1196 GGGATTCGAGAGGATTTTGGACATCTCTGGAAAAATTCGAGGACCCAGTAATCC 1255  
DB 711 GGGATTCGAGAGGATTTTGGACATCTCTGGAAAAATTCGAGGATCCAGTAATCC 770  
QY 1256 TGCCATCATCAGGAGCTGTGGAAATATATATTCGAAGCTTTTGGCAAGAGCTAAAT 1315  
DB 771 TGCCATCATCAGGAGCTGTGGAAATATATATTCGAAGCTTTTGGCAAGAGCTAAAT 830  
QY 1316 TATTCCTCTTATCTGTAATAATCATGCTAGATCTTTGGTTAACTGGCTGCACATATA 1375  
DB 831 TATTCCTCTTATCTGTAATAATCATGCTAGATCTTTGGTTAACTGGCTGCACATATA 890  
QY 1376 CTTTATTAACAGGATTCGGAAACAAAGGCATCTGCGATGTGCTCTCCATGGACCAT 1435  
DB 891 CTTTATTAACAGGATTCGGAAACAAAGGCATCTGCAATGTGCTCTCCATGGACCAT 950  
QY 1436 TTACTCAGCTCCGAGCTGTGTCTTACACCTTTGTTTATAGACAAACAGCTTTTGAG 1495  
DB 951 TTACTCAGCTCCGAGCTGTGTCTTACACCTTTGTTTATAGACAAACAGCTTTTGAG 1010  
QY 1496 CGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTCAATTTTGCAGGATAGTAT 1555  
DB 1011 CGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTCAATTTTGCAGGATAGTAT 1070  
QY 1556 GAGCGAGCTTAATCCCTGAAGATTTGCTGCTCCCTCAGTGGTTAACTTTTGGCTGCAAT 1615  
DB 1071 GAGCGAGCTTAATCCCTGAAGATTTGCTGCTCCCTCAGTGGTTAACTTTTGGCTGCAAT 1130  
QY 1616 CACAATAAGTACCACTGCTCTTCTGCTACACCATCATTCAGAGGAAACATCGCCAGAT 1675  
DB 1131 CACAATAAGTACCACTGCTCTTCTGCTACACCATCATTCAGAGGAAACATCGCCAGAT 1190  
QY 1676 GTGCGAGCTATTAGAGTACCGCTGGAGAGACTCAGTGCAGATCTGCACAAACCCGCT 1735  
DB 1191 GTGCGAGCTATTAGAGTACCACTGGAGGAGACTCAGTGCAGATCTGCACAAACCCACT 1250  
QY 1736 GGNACACTTCTCCCTTTGATCCCTGTGCTGCTGAAGAGGTCA 1778  
DB 1251 GGACACTTTCGTCCCTTTGATCCCTGTGCTGCTGAAGAGGTGA 1293





Hopkins, R.F., Jordan, H., Moore, T., Max, S. J., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Young, A.C., Shevchenko, Y., Sanchez, A., Whitting, M., Touchman, J.W., Green, E.D., Bonfard, G.G., Blakesley, R.W., Grimwood, J., Schmutz, J., Myers, R.M., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalus, D.E., Schnerker, A., Schein, J.E., Jones, S.J. and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

23488257  
22477932  
2 (bases 1 to 2243)  
Strausberg, R.  
Direct Submission  
Submitted (01-AUG-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue procurement: Dr. Jim Lin, University of Iowa  
cDNA Library Preparation: M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>  
Series: IRAC Plate: 125 Row: c Column: 18  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

Location/Qualifiers  
1. .2243  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6403754"  
/tissue type="Brain, mouse 12.5 dpc"  
/clone lib="NIH BMAP\_PD0"  
/lab host="DH10B"  
/note="Vector: pYX-ASC"  
<1. .2243  
/gene="AL023001"  
/note="synonym: Rrn3"  
/db\_xref="LocusID:106298"  
/db\_xref="MGI:2146525"  
<1. .647  
/gene="AL023001"  
/codon\_start=3  
/product="AL023001 protein"  
/protein\_id="AAH55781.1"  
/db\_xref="GI:33585914"  
/db\_xref="LocusID:106298"  
/db\_xref="MGI:2146525"  
/translation="KAFCDVALHGPPYSACQAVFYTVFHHKLLSGNLKQGLQYLQS  
LNFRIVLSQNLKICLPQVNFFAATNKYOLVFVFNERNRNSQMLPVIKSTAGG  
DSVQCTNPLDTFFPDCPKLRSKFIDPIYQIWEDGSABELQEFKKTKKQVVEDE

DDDFLKGEVPOSDFVTGLTPSSFDTHFQSPSSSVGSPVLYIPGOSPLLLTRIYD  
27. 434  
/gene="AL023001"  
/note="Region: RNA polymerase I specific transcription  
initiation factor RRN3. This family consists of several  
eukaryotic proteins which are homologous to the yeast RRN3  
protein. RRN3 is one of the RRN genes specifically  
required for the transcription of rDNA by RNA polymerase I  
(Pol I) in *Saccharomyces cerevisiae*"  
/db\_xref="CDD:pfam05327"

ORIGIN  
Query Match 23.1%; Score 478; DB 10; Length 2243;  
Best Local Similarity 83.9%; Pred. No. 2.5e-120;  
Matches 552; Conservative 0; Mismatches 105; Indels 1; Gaps 1;  
QY 1399 CAAAGGCATTCTGCGATGTTGCTCTCCATGAGCACTTTTACTCAGCCTGCCAAGCTGTG 1458  
Db 1 CAAAGGCCTTTTGTGACGTTGCACTCCATGGACCACTTTTATTTCAGCCTGCCAAGCTGTG 60  
QY 1459 TCTACACCTTTGTTTTTTAGACACACAGCAGCTTTTTCAGCGGAACTCTGAAAGAGGTTTGC 1518  
Db 61 TCTACACCTGTTGTTTTTTAGACACACAGCAGCTTTTTCAGTGGAACCTTGAAGCAAGTCTTAC 120  
QY 1519 AGTATCTTTCAGAGTCTGAATTTTTCAGCGGATAGTGATGAGCCAGTAAATCCCTCGAAGA 1578  
Db 121 AGTATCTTTCAGAGTCTGAATTTTTCAGCGGATAGTGATGAGCCAGTAAATCCCTCGAAGA 180  
QY 1579 TTTGCTGCTCCTCAGTGGTTAACTTTTTCGTCGCAATCAAAATAGTACAGCTCGTCT 1638  
Db 181 TCTGCTGCTGCGCAAGTGGTTAACTTTTTCGTCGCTATCAAAATAATACCAAGTGGTGT 240  
QY 1639 TCTGCTACACCATCATTTGAGAGGAAACAATCGCAGATGCTGCAGTCATAGGAGTACCG 1698  
Db 241 TCTGCTACACCATCATTTGAGAGGAAACAATCGCAGATGCTGCAGTCTCCCGTTATCCGACACCG 300  
QY 1699 CTGAGGAGACTCAGTGCAGATCTGCACAAACCCGTCGACACTTCTTCCCTTTTGATC 1758  
Db 301 CTGTTGGAGACTCCGTCGAGACCTGCACCAACCCAGTGCACACTTCTTCCCTTTTGACC 360  
QY 1759 CTTGTTGCTGAAGAGGTCACAAAGAAATTCATTGATCTCTATTATTCAGGTATGGGAAGA 1818  
Db 361 CTTGTTGCTGAAGAGGTCACAAAGAAATTCATTGATCTCTATTATTCAGATTTGGGAAGATG 420  
QY 1819 TCAGTGTCTGAAGAGTCTACAGGAGTTTCAAGAAACCCATGAAAGAGGACATAGTGGAAAGATG 1878  
Db 421 GGAGTGTCTGAAGAGTCTCAGGAGTTTGAAGAAATCTACTAAAAGCAGGTAGTGGAGATG 480  
QY 1879 AAGATGATGACTTCTTGAAAGCGAGTGCCTCCAGAAATGATACCGTGTATGGGATCACAC 1938  
Db 481 AAGATGATGACTTCTTGAAAGCGAGTGCCTCCAGAGTGCACAGTGCCTTACCTC 540  
QY 1939 CAAAGCTCTTTGACACGCAATTTCCGAAAGTCTTCAAGTGTGGGCTCCCAACCGTGT 1998  
Db 541 CGAGCTCTCTTGTATACCCACTTCCAAAGTCTTCCAGTGTGGGCTCCCTCTCTGTGC 600  
QY 1999 TGTACAT-GCAACCCAGTCCCTCTGACGCGCAAAATTTGTGACTCAGATGTGACAT 2055  
Db 601 TGTATATACAGCCAGTCTCCACTCTCACAAGGATCTATGATTGAGCTGGCAAT 658

RESULT 13  
AX274907  
LOCUS AX274907 1461 bp DNA linear PAT 29-OCT-2001  
DEFINITION Sequence 172 from Patent WO0172777.  
ACCESSION AX274907  
VERSION AX274907.1 GI:16547539  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE

AUTHORS Hillman, J.L., Baughn, M.R., Yue, H., Lal, P., Lu, D.A., Patterson, C., Azinza, Y., Bandman, O., Tang, Y.T., Mathur, P., Shah, P., Au-Young, J., and Reddy, R.

TITLE Transcription factors

JOURNAL Patent: WO 0172777-A 172 04-OCT-2001; Incyte Genomics, Inc. (US)

FEATURES

source

1..1461  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 2646274CB1"

ORIGIN

Query Match 20.3%; Score 420.8; DB 6; Length 1461;  
Best Local Similarity 97.3%; Pred. No. 1.5e-104;  
Matches 428; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1188 TTCAAAATGGGATTCGCAGAGGCATTTTTGGACATCTCTGAAAAAATTCGAGACCCA 1247  
Db |||

QY 530 TTTAATATGGGATTCGCAGAGGCATTTTTGGACATCTCTTGGAAAACTTCGAGATCCA 589  
Db |||

QY 1248 AGTAATCTGCCATCATCAGGAGGCTGCTGAAATATATATTGGAAGCTTTTGGCAAGA 1307  
Db |||

QY 590 AGTAATCTGCCATCATCAGGAGGCTGCTGAAATATATATTGGAAGCTTTTGGCAAGA 649  
Db |||

QY 1308 GCTAAATTTATCTCTTATTACTGTAATAATCATGCCAGATCTTTTGGTAACTGGCTG 1367  
Db |||

QY 650 GCTAAATTTATCTCTTATTACTGTAATAACCATGCCAGATCTTTTGGTAACTGGCTG 709  
Db |||

QY 1368 CACATATACCTTAATPACAGGATTCGGACAAAGGATCTCTGAGTGTGCTCTCCAT 1427  
Db |||

QY 710 CACATATACCTTAATPACAGGATTCGGACAAAGGATCTCTGAGTGTGCTCTCCAT 769  
Db |||

QY 1428 GGACCATTTTACTAGCCTGCCAGCTGTGTCTACACCTTTGTTTATAGACACAGCAG 1487  
Db |||

QY 770 GGACCATTTTACTAGCCTGCCAGCTGTGTCTACACCTTTGTTTATAGACACAGCAG 829  
Db |||

QY 1488 CTTTGGAGCGGAACCTGAAGAAGGTTTGAGTATCTTCAGATCTCGAATTTTGGAGCG 1547  
Db |||

QY 830 CTTTGGAGCGGAACCTGAAGAAGGTTTGAGTATCTTCAGATCTCGAATTTTGGAGCG 889  
Db |||

QY 1548 ATAGTGATGAGCGAGTAAATCCCTGAGATTTGCTGAGTATCTTCAGATCTCGAATTTT 1607  
Db |||

QY 890 ATAGTGATGAGCGAGTAAATCCCTGAGATTTGCTGAGTATCTTCAGATCTCGAATTTT 949  
Db |||

QY 1608 GCTGCAATCACAATAAGTA 1627  
Db |||

QY 950 GCTGCAATCACAATAAGTA 969  
Db |||

RESULT 14

BC006441

LOCUS

DEFINITION Homo sapiens similar to RNA polymerase I transcription factor RN3, mRNA (cDNA clone IMAGE:3351791), partial cds.

ACCESSION BC006441

VERSION BC006441.1 GI:13623642

KEYWORDS

SOURCE

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 2627)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gumaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Halyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 2627)

Strausberg, R.

Direct Submission

Submitted (09-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada

[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Khyu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smal, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 18 Row: d Column: 24.

Location/Qualifiers

1..2627

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3351791"

/tissue type="Eye, retinoblastoma"

/clone lib="NIH MGC 16"

/lab host="DH10B-R"

/note="Vector: pOTB7"

ORIGIN

Query Match 20.3%; Score 420.8; DB 9; Length 2627;  
Best Local Similarity 97.3%; Pred. No. 1.7e-104;  
Matches 428; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1188 TTCAAAATGGGATTCGCAGAGGCATTTTTGGACATCTCTGAAAAAATTCGAGACCCA 1247  
Db |||

QY 1692 TTTAATATGGGATTCGCAGAGGCATTTTTGGACATCTCTTGGAAAACTTCGAGATCCA 1751  
Db |||

QY 1248 AGTAATCTGCCATCATCAGGAGGCTGCTGAAATATATATTGGAAGCTTTTGGCAAGA 1307  
Db |||

QY 1752 AGTAATCTGCCATCATCAGGAGGCTGCTGAAATATATATTGGAAGCTTTTGGCAAGA 1811  
Db |||

QY 1308 GCTAAATTTATCTCTTATTACTGTAATAATCATGCCAGATCTTTTGGTAACTGGCTG 1367  
Db |||

QY 1812 GCTAAATTTATCTCTTATTACTGTAATAACCATGCCAGATCTTTTGGTAACTGGCTG 1871  
Db |||

QY 1368 CACATATACCTTAATPACAGGATTCGGACAAAGGATCTCTGAGTGTGCTCTCCAT 1427  
Db |||

QY 1872 CACATATACCTTAATPACAGGATTCGGACAAAGGATCTCTGAGTGTGCTCTCCAT 1931  
Db |||

QY 1428 GGACCAATTTTACTGACGCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAGCAG 1487  
DB 1932 GGACCAATTTTACTGACGCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAGCAG 1991  
QY 1488 CTTTGTAGCGGAAACCTGAAAGAGCTTTGACAGTATCTTCAGAGTCTGAATTTTTCAGCGG 1547  
DB 1992 CTTTGTAGCGGAAACCTGAAAGAGCTTTGACAGTATCTTCAGAGTCTGAATTTTTCAGCGG 2051  
QY 1548 ATAGTGATGAGCAGCACTAAATCCCTGGAAGATTTGCTGCCCTCAGTGGTTAACTTTTT 1607  
DB 2052 ATAGTGATGAGCAGCACTAAATCCCTGGAAGATTTGCTGCCCTCAGTGGTTAACTTTTT 2111  
QY 1608 GCTGCAATCACAATAAGTA 1627  
DB 2112 GCTGCAATCACAATAAGTA 2131

RESULT 15  
AK126166  
LOCUS AK126166 3645 bp mRNA linear PRI 09-SEP-2003  
DEFINITION Homo sapiens cDNA FLJ41178 fis, clone THYMU2037233, highly similar  
to Homo sapiens RNA polymerase I transcription factor RRN3 (RRN3).  
ACCESSION AK126166  
VERSION AK126166.1 GI:34532564  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1. Oshima A., Takahashi-Fujii A., Tanase T., Inose N., Takeuchi K.,  
Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,  
Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,  
Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,  
Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H.,  
Kanda K., Wagatsuma M., Murakawa K., Kanehori K., Sugiyama A.,  
Kawakami B., Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K.  
and Isogai F.  
NEDO human cDNA sequencing project

TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3645)  
AUTHORS Isogai T. and Yamamoto J.  
TITLE Direct Submission  
JOURNAL Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 252-0818, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.

FEATURES  
Source  
1..3645  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="THYMU2037233"  
/tissue\_type="thymus"  
/clone\_lib="THYMU2"  
/note="cloning vector: pME18SFL3"

ORIGIN  
Query Match 20.3%; Score 420.8; DB 9; Length 3645;  
Best Local Similarity 97.3%; Pred. No. 1.8e-104;  
Matches 428; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1188 TTCAATTTGGATTCGAGGAGGATTTTGGACATCTCTGGAAAAATTCGAGGCCCA 1247  
DB 2771 TTTAATATGGATTCGAGGAGGATTTTGGACATCTTTGGAAAACTTCGAGGATCCA 2830

QY 1248 AGTAATCCTGCCATCATCAGCAGGCTGCTGGAATTTATATTGGAAGCTTTTGGCAAGA 1307  
DB 2831 AGTAATCCTGCCATCATCAGCAGGCTGCTGGAATTTATATTGGAAGCTTTTGGCAAGA 2890  
QY 1308 GCTAAATTTATTTCCCTCTTTATTACTGTAAATCATGSCCTAGATCTTTTGGTTAACTGGCTG 1367  
DB 2891 GCTAAATTTATTTCCCTCTTTATTACTGTAAACCATGSCCTAGATCTTTTGGTTAACTGGCTG 2950  
QY 1368 CACATATACCTTATATAACAGGATTCGGGAACAAAGGCATTTCTGCGATGTTGCTCTCCAT 1427  
DB 2951 CACATATACCTTATATAACAGGATTCGGGAACAAAGGCATTTCTGCGATGTTGCTCTCCAT 3010  
QY 1428 GGACCAATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAGCAG 1487  
DB 3011 GGACCAATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAGCAG 3070  
QY 1488 CTTTGTAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTTCAGCGG 1547  
DB 3071 CTTTGTAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAATTTTTCAGCGG 3130  
QY 1548 ATAGTGATGAGCAGCACTAAATCCCTGGAAGATTTGCTGCCCTCAGTGGTTAACTTTTT 1607  
DB 3131 ATAGTGATGAGCAGCACTAAATCCCTGGAAGATTTGCTGCCCTCAGTGGTTAACTTTTT 3190  
QY 1608 GCTGCAATCACAATAAGTA 1627  
DB 3191 GCTGCAATCACAATAAGTA 3210

Search completed: July 18, 2004, 04:52:49  
Job time : 8122 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 01:13:34 ; Search time 816 Seconds  
(without alignments)  
10766.267 Million cell updates/sec

Title: US-09-932-678-1

Perfect score: 2068

Sequence: 1 acagaggctgtggtggaag.....tgacatttgggtattcccat 2068

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_29Jan04.\*

- 1: geneseq1980s.\*
- 2: geneseq1990s.\*
- 3: geneseq2000s.\*
- 4: geneseq2001as.\*
- 5: geneseq2001bs.\*
- 6: geneseq2002s.\*
- 7: geneseq2003as.\*
- 8: geneseq2003bs.\*
- 9: geneseq2003cs.\*
- 10: geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2068	100.0	2068	6	ABV72159 Nucleotid
2	2018.4	97.6	2040	3	AAA98384 Human RNA
3	1457.6	70.5	1770	6	ABs78724 DNA encod
4	1253.6	60.6	1582	9	ADD29697 Human tum
5	1228.4	59.4	1418	3	AAC76937 Human ORF
6	1010.4	48.9	1498	4	AAS44584 Human ful
7	894	43.2	1423	4	AAS01562 Human sec
8	639.4	30.9	2410	5	AAS92255 DNA encod
9	568.6	27.5	776	5	AAS92253 DNA encod
10	567	27.4	2493	6	ABs64829 Human tum
11	567	27.4	2493	7	ADA52976 Human cod
12	550.8	26.6	1260	5	AAS87643 DNA encod
13	477.8	23.1	655	5	AAS80955 DNA encod
14	466.2	22.5	3169	5	AAS87641 DNA encod
15	420.8	20.3	1461	5	ABa83038 Human tra
16	420	20.3	437	2	AAX51663 Human sec
17	419.2	20.3	1989	6	ABK47950 Human tra
18	398	19.2	560	5	AAS91991 DNA encod
19	390.6	18.9	3175	5	AAS92254 DNA encod
20	387	18.7	585	5	AAS92248 DNA encod
21	315	15.2	1017	5	AAS87636 DNA encod
22	292	14.1	341	2	AAV90248 EST clone
23	286.6	13.9	498	5	AAS92247 DNA encod

24	282.2	13.6	396	5	AAS91986 DNA encod
25	282.2	13.6	396	5	AAS80954 DNA encod
26	282.2	13.6	396	5	AAS92246 DNA encod
27	282.2	13.6	396	5	AAS87635 DNA encod
28	279	13.5	2991	5	AAS91992 DNA encod
29	279	13.5	2991	5	AAS73660 DNA encod
30	247.4	12.0	349	3	AAC77533 Human ORF
31	233.4	11.3	667	5	AAS92251 DNA encod
32	230	11.1	2103	3	AAC77798 Human can
33	208	10.1	34269	4	AAK68677 Human imm
34	208	10.1	34269	4	AAK85168 Human imm
35	203.6	9.8	213	3	AAC19091 Human sec
36	192	9.3	17803	4	AAK68676 Human imm
37	188.8	9.1	9259	4	AAK68674 Human imm
38	188.8	9.1	23815	4	AAK85169 Human imm
39	188.8	9.1	23815	4	AAK68678 Human imm
40	163.4	7.9	627	5	AAS87640 DNA encod
41	156	7.5	557	6	ABV87866 Human col
42	144.2	7.0	3410	4	AAK85167 Human imm
43	139.4	6.7	1326	5	AAS91907 DNA encod
44	139.4	6.7	3042	5	AAS70319 DNA encod
45	138.6	6.7	807	5	AAS92252 DNA encod

#### ALIGNMENTS

RESULT 1  
ABV72159  
ID ABV72159 standard; DNA; 2068 BP.  
XX  
AC ABV72159;  
XX  
DT 05-DEC-2002 (first entry)  
XX  
DE Nucleotide sequence of human RRN3.  
XX  
KW Human; RRN3; RNA polymerase I transcription factor;  
KW hypoproliferative disease; hyperproliferative disease; cancer;  
KW malignancy; hyperplasia; metaplasia; dysplasia; benign tumour;  
KW hyperproliferative disorder; benign dysproliferative disorder;  
KW autoimmune disease; cardiac disease; degenerative disorder;  
KW growth deficiency; hypoproliferative disorder; physical trauma; lesion;  
KW wound; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 69..2024  
FT /\*tag= a  
FT /product= "RRN3"  
US2002090706-A1.  
11-JUL-2002.  
16-AUG-2001; 2001US-00932678.  
16-AUG-2000; 2000US-0225893P.  
(REED/) REEDER R H.  
(MOOR/) MOOREFIELD B.  
(GREE/) GREENE E A.  
Reeder RH, Moorefield B, Greene EA;  
WPI; 2002-681729/73.  
P-PSDB; ABB78308.  
Novel RRN3 polypeptide, an eukaryotic RNA polymerase I transcription factor and polynucleotide encoding it useful for diagnosing, treating hyper and hypoproliferative diseases in mammals and to identify modulators.

XX PS Claim 6; Page 25-26; 31pp; English.

XX CC The present sequence encodes a human RN3 polypeptide. RN3 is an

CC eukaryotic RNA polymerase I transcription factor. RN3 is useful for

CC screening for hypo or hyperproliferative diseases, including cancer,

CC malignancy, hyperplasia, metaplasia, dysplasia, benign tumour,

CC hyperproliferative disorder, benign dysproliferative disorder, autoimmune

CC disease or cardiac disease. It is also used to treat diseases involving

CC decreased cell proliferation, including degenerative disorders, growth

CC deficiencies, hypoproliferative disorders, physical trauma, lesions and

CC wounds. RN3 polypeptide, nucleic acid or antibody are also useful in

CC diagnostics, to detect, prognose, diagnose or monitor various diseases.

CC RN3 nucleic acids are useful to identify other mammalian genes that

CC encode RN3-like molecules, to screen for mutations in a RN3 gene that

CC are associated with certain diseases. The polypeptide is useful as an

CC immunogen to generate antibodies which immunospecifically bind RN3

XX polypeptides

SQ Sequence 2068 BP; 573 A; 437 C; 485 G; 573 T; 0 U; 0 Other;

Query Match 100.0%; Score 2068; DB 6; Length 2068;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2068; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGAGGCTGTGGTGAAGAGGCTGGGATCGGCTGAGCGGAGCGGTCGGTGTAGT 60

DB 1 ACAGAGGCTGTGGTGAAGAGGCTGGGATCGGCTGAGCGGAGCGGTCGGTGTAGT 60

QY 61 TCGGCCCAATGCGGCAACCGCTGCTTACACGCGTTTGGCGGAGATGGCGGCTTCGT 120

DB 61 TCGGCCCAATGCGGCAACCGCTGCTTACACGCGTTTGGCGGAGATGGCGGCTTCGT 120

QY 121 CCTCTGCACTTAAGAGCTGGGCGCTCGAGGACTGGGATTTCAATATGCGTGCATTAG 180

DB 121 CCTCTGCACTTAAGAGCTGGGCGCTCGAGGACTGGGATTTCAATATGCGTGCATTAG 180

QY 181 AGAATGACTTTTTCAANTCTCCCAAGAAAACCTGTTGGTGGTGAACCTGTGACAG 240

DB 181 AGAATGACTTTTTCAANTCTCCCAAGAAAACCTGTTGGTGGTGAACCTGTGACAG 240

QY 241 AAGTCTTGCTGAAGTACAAAAGGGTGAAACAAATGACTTTTGAGTTGTTGAAGAACGAGC 300

DB 241 AAGTCTTGCTGAAGTACAAAAGGGTGAAACAAATGACTTTTGAGTTGTTGAAGAACGAGC 300

QY 301 TGTATGATCCAGACATAAAGGATGACAGATCATCAACTGGCTGCTAGAAATTCGGTTCCT 360

DB 301 TGTATGATCCAGACATAAAGGATGACAGATCATCAACTGGCTGCTAGAAATTCGGTTCCT 360

QY 361 CTATCATGTACTTGACAAAAGACTTTGACCAACTATCAGTATATTAAGATTCGCTT 420

DB 361 CTATCATGTACTTGACAAAAGACTTTGACCAACTATCAGTATATTAAGATTCGCTT 420

QY 421 GGTGTAATGAACTCAAAAGAGTGTGAAGATTTGGCTTTCTTGGTAAATCTTGTAT 480

DB 421 GGTGTAATGAACTCAAAAGAGTGTGAAGATTTGGCTTTCTTGGTAAATCTTGTAT 480

QY 481 CAGCAGAGCTGTTTCTCAGACCGTGTCTAGAGATGATGCTTCCCATTTTGTGCTC 540

DB 481 CAGCAGAGCTGTTTCTCAGACCGTGTCTAGAGATGATGCTTCCCATTTTGTGCTC 540

QY 541 CCGAGTGTATCAATTAAGGAAGCGATGTAGATGTTTCAGATTTCTGATGAAGATGATA 600

DB 541 CCGAGTGTATCAATTAAGGAAGCGATGTAGATGTTTCAGATTTCTGATGAAGATGATA 600

QY 601 ATCTTCTGCAAAATTTTGACACATGTCCAGAGCCTTGGAAAATTAATAGCAAGATATGTAC 660

DB 601 ATCTTCTGCAAAATTTTGACACATGTCCAGAGCCTTGGAAAATTAATAGCAAGATATGTAC 660

QY 661 CATCGACACCGTGTGTTCTCATGCCAATACCTGGTGGAAAATTTCCATTTGTTTCGAAAT 720

DB 661 CATCGACACCGTGTGTTCTCATGCCAATACCTGGTGGAAAATTTCCATTTGTTTCGAAAT 720

QY 721 CAGAGAGAACACTCGAATGTTACGTTTCATAACTTACTAAGGATTAGTGTATATTTTCCAA 780

DB 721 CAGAGAGAACACTCGAATGTTACGTTTCATAACTTACTAAGGATTAGTGTATATTTTCCAA 780

QY 781 CTTGAGGCATGAATTCCTGAGCTTATATTTGAAAACTACTCAAGTTGATGTGAATG 840

DB 781 CTTGAGGCATGAATTCCTGAGCTTATATTTGAAAACTACTCAAGTTGATGTGAATG 840

QY 841 CATCCCGCAGGATTTGAAGATCTGAAGAAACAGCAACTCAAACTTTGTTGGTGGACAG 900

DB 841 CATCCCGCAGGATTTGAAGATCTGAAGAAACAGCAACTCAAACTTTGTTGGTGGACAG 900

QY 901 ATTCCACGGAAGGATTTGTTTAATATGAGATGAAGAACTGAACATGAACAAAGG 960

DB 901 ATTCCACGGAAGGATTTGTTTAATATGAGATGAAGAACTGAACATGAACAAAGG 960

QY 961 CTGTCCTGAACCGGCTCGACGATGGTGCATCTCTGTAGCCGAGCGCTGGACATCTCTGA 1020

DB 961 CTGTCCTGAACCGGCTCGACGATGGTGCATCTCTGTAGCCGAGCGCTGGACATCTCTGA 1020

QY 1021 TGTCTTTGTTTGTCTTACATGAAGGATGCTGCTATGTATGATGTTAAGTTGATAACG 1080

DB 1021 TGTCTTTGTTTGTCTTACATGAAGGATGCTGCTATGTATGATGTTAAGTTGATAACG 1080

QY 1081 GCAAAACAAAGGATCTATATCGGACCTGATAAACATCTTTGACAAACTCCTGTTGCCCA 1140

DB 1081 GCAAAACAAAGGATCTATATCGGACCTGATAAACATCTTTGACAAACTCCTGTTGCCCA 1140

QY 1141 CCCATGCTCTGCGCATGTACAGTGTTCATGTTTTTACCTCTGTAGTTTCAATTTGGGAT 1200

DB 1141 CCCATGCTCTGCGCATGTACAGTGTTCATGTTTTTACCTCTGTAGTTTCAATTTGGGAT 1200

QY 1201 TCGCAGAGGCACTTTTGGAAACATCTCGAAAAAATTTGAGGACCCCAAGTAATCCTGCGCA 1260

DB 1201 TCGCAGAGGCACTTTTGGAAACATCTCGAAAAAATTTGAGGACCCCAAGTAATCCTGCGCA 1260

QY 1261 TCATCAGGCGAGGCTGCTGAAAATTTATATTTGGAAGCTTTTGGCAAGAGCTAAATTTATTC 1320

DB 1261 TCATCAGGCGAGGCTGCTGAAAATTTATATTTGGAAGCTTTTGGCAAGAGCTAAATTTATTC 1320

QY 1321 CTCCTATTACTGTAAAAATCATGCTTAGATCTTTTGGTTAACTGGCTGACATATACCTTA 1380

DB 1321 CTCCTATTACTGTAAAAATCATGCTTAGATCTTTTGGTTAACTGGCTGACATATACCTTA 1380

QY 1381 ATAAACAGGATTTGGGAAACAAAGGATTTCTGCGATCTGCTCTCCATGGACCAATTTTACT 1440

DB 1381 ATAAACAGGATTTGGGAAACAAAGGATTTCTGCGATCTGCTCTCCATGGACCAATTTTACT 1440

QY 1441 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTGAACAAGAGCTTTTGGAGCGGAA 1500

DB 1441 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTGAACAAGAGCTTTTGGAGCGGAA 1500

QY 1501 ACTGAAAGAGGTTTGCAGTATCTTCAGAGTGTGAATTTTGGCGGATGATGATGAGCC 1560

DB 1501 ACTGAAAGAGGTTTGCAGTATCTTCAGAGTGTGAATTTTGGCGGATGATGATGAGCC 1560

QY 1561 AGCTAAATCCCTGGAAGATTTGCTGCTCCTCAGTGTGTTAACTTTTGTGCAATCACA 1620

DB 1561 AGCTAAATCCCTGGAAGATTTGCTGCTCCTCAGTGTGTTAACTTTTGTGCAATCACA 1620

QY 1621 ATAAGTACAGCTGCTGCTTCTGCTACACCATATTGAGAGGAAATTCGCGAGATGCTG 1680

DB 1621 ATAAGTACAGCTGCTGCTTCTGCTACACCATATTGAGAGGAAATTCGCGAGATGCTG 1680

QY 1681 CAGTCAATTAGGAGTACCGCTGGAGGAGCTCAGTGCAGATCTGCACAAACCCCTGGACA 1740

DB 1681 CAGTCAATTAGGAGTACCGCTGGAGGAGCTCAGTGCAGATCTGCACAAACCCCTGGACA 1740

QY 1741 CCTTCTTCCCTTTTGTATCCTGCTGAAGAGGTCAAAAGAAATTCATTGATCCTTATTT 1800

DB 1741 CCTTCTTCCCTTTTGTATCCTGCTGAAGAGGTCAAAAGAAATTCATTGATCCTTATTT 1800

QY 1801 ATCAGGTATGGGAAGACATGAGTGTCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAA 1860

Db 1801 ATCAGTATGGGAAGACATGAGTGTCTGAGAGCTACAGAGTTCAGAAACCCATGAAAA 1860  
QY 1861 AGGACATAGTGGAGATGAAGATGATGATCTTCTGAAAGCGGAAGTGCCTCCAGAAATGATA 1920  
Db 1861 AGGACATAGTGGAGATGAAGATGATGATCTTCTGAAAGCGGAAGTGCCTCCAGAAATGATA 1920  
QY 1921 CCGTGAATGGGATCACACCAAGCTCTTTCGACACGCAATTCGGAAGTCTTCAAGTAGTG 1980  
Db 1921 CCGTGAATGGGATCACACCAAGCTCTTTCGACACGCAATTCGGAAGTCTTCAAGTAGTG 1980  
QY 1981 TGGGTCCGCCACCGTGTGTATGATCAACCCAGTCCCTCTGACGGCAGAAATTTGTGA 2040  
Db 1981 TGGGTCCGCCACCGTGTGTATGATCAACCCAGTCCCTCTGACGGCAGAAATTTGTGA 2040  
QY 2041 CTGAGATGTGACATTTGGGATTCCTCAT 2068  
Db 2041 CTGAGATGTGACATTTGGGATTCCTCAT 2068

RESULT 2  
AAA98384  
ID AAA98384 standard; cDNA; 2040 BP.  
AC AAA98384;  
DT 08-FEB-2001 (first entry)  
DE Human RNA polymerase I transcription factor TIF-1A cDNA.  
XX  
KW RNA polymerase I transcription factor TIF-1A; antitumor; treatment;  
antiproliferative; cell proliferation; cancer; tissue regeneration; ss.  
XX  
OS Homo sapiens.  
XX  
EN WO200055316-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000WO-DE000767.  
XX  
FR 17-MAR-1999; 99DE-01011992.  
XX  
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
XX  
PI Grummt I, Vingron M;  
XX  
DR WPI; 2000-587527/55.  
DR P-PSDB; AAB10936.  
XX  
XX  
XX

New DNA encoding the transcription factor TIF-1A, useful for preventing or treating diseases associated with abnormal cell proliferation, particularly tumors.  
Claim 1; Fig 2; 38pp; German.  
This invention describes a novel DNA sequence (I) that encodes the RNA polymerase I transcription factor TIF-1A which has antitumor, antiproliferative and proliferation-inducing activity. The invention also describes (1) DNA (Ia) encoding a protein (II) with the biological activity of TIF-1A; (2) a ribozyme (R) corresponding to (I) or (Ia) and able to bind specifically to, and cleave, its transcribed RNA so as to reduce or inhibit synthesis of the corresponding protein; (3) an antisense RNA (AS) with binding properties similar to R; (4) an expression vector that contains (I), (Ia) or sequences that encode R or AS; (5) host cells containing the vectors of (4); (6) TIF-1A or (II) encoded by (I) or (Ia); (7) preparation of TIF-1A or (II) by culturing cells of (6); (8) ligands that bind to TIF-1A or (II); (9) antagonists that weaken or block the activity of TIF-1A or (II); (10) a diagnostic method for detecting abnormal TIF-1A expression; and (11) a kit for carrying out the method in (10). (I), and similar sequences that encode proteins with equivalent activity, expression vectors containing them, or their expression products are used to treat or prevent disorders

CC associated with reduced cellular proliferation, to stimulate cellular proliferation, and to promote tissue regeneration, e.g. after injury or radiation therapy. Ribozymes, antisense sequences directed against (I), CC also ligands and antagonists of TIF-1A, are used to treat or prevent CC disorders associated with excessive cellular proliferation and to inhibit CC proliferation, especially in treatment of cancers. (I) and specific CC ligands for TIF-1A (particularly antibodies (Ab)) are also useful for CC diagnosis of altered TIF-1A expression by (in)direct determination of the CC concentration, length and/or sequence of TIF-1A or its mRNA, e.g. for CC detecting mutations. Ab can also be used for immunoprecipitation of TIF-1A and for isolation of related sequences from cDNA expression libraries. CC (I) allows recombinant production of TIF-1A in sufficient quantities for CC therapeutic use  
XX

SQ Sequence 2040 BP; 563 A; 433 C; 474 G; 568 T; 0 U; 2 Other;

Query Match 97.6%; Score 2018.4; DB 3; Length 2040;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2019; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 47 GCGGTCGCGTTAGTTCCGCCCAATGGCGSCACCGCTGCTTCACACGCTTTGCCGGGAGA 106  
Db 18 GCGGTCGCGTTAGTTCCGCCCAATGGCGSCACCGCTGCTTCACACGCTTTGCCGGGAGA 77  
QY 107 TCGGCGCGCTTCGTCCTCTGCGAGTTAAGAGCTGGCGCGCTCGAGGACTGGGATTTCAAA 166  
Db 78 TCGGCGCGCTTCGTCCTCTGCGAGTTAAGAGCTGGCGCGCTCGAGGACTGGGATTTCAAA 137  
QY 157 TATGGTGCATAGAGAAATGACTTTTCAATTCCTCCCAAGAAAACCTGTTGGGTTGG 226  
Db 138 TATGGTGCATAGAGAAATGACTTTTCAATTCCTCCCAAGAAAACCTGTTGGGTTGG 197  
QY 227 TGGAACTGTGACAGAGTCTTCTGCTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTT 286  
Db 198 TGGAACTGTGACAGAGTCTTCTGCTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTT 257  
QY 287 GTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGATGACAGATCATCAACTGCTGCT 346  
Db 258 GTTGAAGAACCCAGCTGTTAGATCCAGACATAAAGATGACAGATCATCAACTGCTGCT 317  
QY 347 AGAATTCGTTCTTCTATCATGTAATGACAAAGACTTTGAGCAACTTATCAGTATTAT 406  
Db 318 AGAATTCGTTCTTCTATCATGTAATGACAAAGACTTTGAGCAACTTATCAGTATTAT 377  
QY 407 ATTAAGATTCGTTGTTGAATGAACTGAAACAGTAGTGAAGACTTTGGCTTTCT 466  
Db 378 ATTAAGATTCGTTGTTGAATGAACTGAAACAGTAGTGAAGACTTTGGCTTTCT 437  
QY 467 TGGTAATCTTGTATCAGCAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 526  
Db 438 TGGTAATCTTGTATCAGCAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 497  
QY 527 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAGATTCTGA 586  
Db 498 CCATTTTGTGCTCCCGAGTGATCATTAAGGAAGCGGATGTAGATGTTTCAGATTCTGA 557  
QY 587 TGATGAAGATGATAATCTTCTGCAATTTTGACATGTTCAGAGACCTTGCAATAAT 646  
Db 558 TGATGAAGATGATAATCTTCTGCAATTTTGACATGTTCAGAGACCTTGCAATAAT 617  
QY 647 AGCAAGATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGAAAAATTTCC 706  
Db 618 AGCAAGATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGAAAAATTTCC 677  
QY 707 ATTTGTTGCAAAATCAGAGAAACATCGGAATTTACGTTTACATACTAAGGATTAG 766  
Db 678 ATTTGTTGCAAAATCAGAGAAACATCGGAATTTACGTTTACATACTAAGGATTAG 737  
QY 767 TGTATATTTTCCAACTTCAGGCAATCAATCTCGAGCTTATTTGAAAAACTACTCAA 826  
Db 738 TGTATATTTTCCAACTTCAGGCAATCAATCTCGAGCTTATTTGAAAAACTACTCAA 797  
QY 827 GTTGGATGTGAATGCATCCCGGAGGTATTGAAGATGCTGAAGAAACAGCAACTCAAC 886

Db 798 GTTGGATGTGAATGCATCCCGCAGGGTATTGAAGATCTGAAGAAACAGCAACTCAAAAC 857  
Qy 887 TTGTGGTGGGACAGATCCACGAGAGGATGTGTTAATATGATGAAGATGAAGAACTGA 946  
Db 858 TTGTGGTGGGACAGATCCACGAGAGGATGTGTTAATATGATGAAGATGAAGAACTGA 917  
Qy 947 ACATGAACAAAGGCTGGTCCCTGAAACGGCTCGACAGATGGTGCATCTGTAGCCGAGCG 1006  
Db 918 ACATGAACAAAGGCTGGTCCCTGAAACGGCTCGACAGATGGTGCATCTGTAGCCGAGCG 977  
Qy 1007 CCTGACATCTGTATGTCTTTGGTTTGTCTTACATGAAGATGCTCTATGTAGATGG 1066  
Db 978 CCTGACATCTGTATGTCTTTGGTTTGTCTTACATGAAGATGCTCTATGTAGATGG 1037  
Qy 1067 TAAAGTTGATTAACGCAACAAAGGATCTATATCGGACCTGTATAAACATCTTTGACAA 1126  
Db 1038 TAAAGTTGATTAACGCAACAAAGGATCTATATCGGACCTGTATAAACATCTTTGACAA 1097  
Qy 1127 ACTCTGTGTGCCACCCATGCTCTGCAATGTACAGTTTTTTCATGTTTACCTCTGTAG 1186  
Db 1098 ACTCTGTGTGCCACCCATGCTCTGCAATGTACAGTTTTTTCATGTTTACCTCTGTAG 1157  
Qy 1187 TTTGCAATGCGATTCGACAGGCAATTTTGGACATCTCTGAAAAAATTCAGGACCC 1246  
Db 1158 TTTGCAATGCGATTCGACAGGCAATTTTGGACATCTCTGAAAAAATTCAGGACCC 1217  
Qy 1247 AAGTAATCTGCAATCATCAGGAGGCTGCTGAAATTAATTTGGAAGCTTTTGGCAAG 1306  
Db 1218 AAGTAATCTGCAATCATCAGGAGGCTGCTGAAATTAATTTGGAAGCTTTTGGCAAG 1277  
Qy 1307 AGCTAAATTTATTCCTTTATCTGTAATAATCATGCCCTAGATCTTTTGGTTAACTGGCT 1366  
Db 1278 AGCTAAATTTATTCCTTTATCTGTAATAATCATGCCCTAGATCTTTTGGTTAACTGGCT 1337  
Qy 1367 GCACATATACCTTAATAACAGATTCGGGAACAAAGGATTCCTGGATGTTGCTCTCCA 1426  
Db 1338 GCACATATACCTTAATAACAGATTCGGGAACAAAGGATTCCTGGATGTTGCTCTCCA 1397  
Qy 1427 TGGACCAATTTACTCAGCCTGCCAAGCTGTGTCTACACCTTTGTTTGTAGACACAAGCA 1486  
Db 1398 TGGACCAATTTACTCAGCCTGCCAAGCTGTGTCTACACCTTTGTTTGTAGACACAAGCA 1457  
Qy 1487 GCTTTTGGCGGAAAACCTGAAAGAGGTTTGGAGTATCTTCAGAGTCTGAAATTTGAGCG 1546  
Db 1458 GCTTTTGGCGGAAAACCTGAAAGAGGTTTGGAGTATCTTCAGAGTCTGAAATTTGAGCG 1517  
Qy 1547 GATAGTCATGAGCCAGCTAAATCCCTGAAAGATTTGGCTGCCCTCAGTGGTTAACTTTT 1606  
Db 1518 GATAGTCATGAGCCAGCTAAATCCCTGAAAGATTTGGCTGCCCTCAGTGGTTAACTTTT 1577  
Qy 1607 TGTGCAATCACAAATAAGTACCAGTCTGTTCTTCTGTACACCATCATTCAGAGGAAACAA 1666  
Db 1578 TGTGCAATCACAAATAAGTACCAGTCTGTTCTTCTGTACACCATCATTCAGAGGAAACAA 1637  
Qy 1667 TGCCAGATCTGCGAGTCAATAGGAGTACCGCTGGAGGAGTCACTGAGTCTGCAAC 1726  
Db 1638 TGCCAGATCTGCGAGTCAATAGGAGTACCGCTGGAGGAGTCACTGAGTCTGCAAC 1697  
Qy 1727 AAACCCGCTGGACACCTTTTCCCTTTGATCCCTGTGCTGAAGAGGTCAAAGAAATT 1786  
Db 1698 AAACCCGCTGGACACCTTTTCCCTTTGATCCCTGTGCTGAAGAGGTCAAAGAAATT 1757  
Qy 1787 CATGTATCTTATTTATCAGGTATGGGAGACATGAGTCTGTAAGAGCTACAGAGTTCAA 1846  
Db 1758 CATGTATCTTATTTATCAGGTATGGGAGACATGAGTCTGTAAGAGCTACAGAGTTCAA 1817  
Qy 1847 GAAACCCATGAAGAAGACATAGTGGAGATGAAGATGATGATCTTTCTGAAAGCGAAGT 1906  
Db 1818 GAAACCCATGAAGAAGACATAGTGGAGATGAAGATGATGATCTTTCTGAAAGCGAAGT 1877  
Qy 1907 GCCCCAGATGATACCGTGTATTGGGATCACACCAAGCTCTTTTGACACGCAATTTCCGAAG 1966

Db 1878 GCCCCAGAAATGATACCGTGATTGGATCACACCAAGCTCTCTTGACACGCAATTTCCGAAG 1937  
Qy 1967 TCCTTTCAAGTAGTGTGGGCTCCCAACCGTGTGTACATGCAACCCAGTCCCTCTGAGC 2026  
Db 1938 TCCTTTCAAGTAGTGTGGGCTCCCAACCGTGTGTACATGCAACCCAGTCCCTCTGAGC 1997  
Qy 2027 GCAGAAATTTCTGACTGAGATGTGACATTTGGGATTTCCCCAT 2068  
Db 1998 GCAGAAATTTCTGACTGAGATGTGACATTTGGGATTTCCCCAT 2039  
RESULT 3  
ABS78724  
ID ABS78724 standard; DNA; 1770 BP.  
XX AC ABS78724;  
XX DT  
XX 16-DEC-2002 (first entry)  
XX DNA encoding human NOVX3 protein.  
XX Human; NOVX; human disease; NOVX-associated disorder; cancer; addiction;  
KW Hodgkin disease; Von Hippel-Lindau syndrome; Alzheimer's disease; stroke;  
KW tuberosus sclerosis; hypercalcaemia; Parkinson's disease; depression;  
KW Huntington's disease; cerebellar palsy; epilepsy; Lesch-Nyhan syndrome;  
KW multiple sclerosis; ataxia-telangiectasia; leukodystrophy; anxiety; pain;  
KW obesity; Crohn's disease; osteoporosis; inflammatory bowel disease;  
KW infertility; inflammatory bowel disease; atherosclerosis; hypertension;  
KW scleroderma; haemophilia; diabetes; pancreatitis; autoimmune disease;  
KW asthma; arthritis; immunodeficiency; HIV; viral infection; neurogenesis;  
KW bacterial infection; parasitic infection; graft-versus-host disease;  
KW cell differentiation; cell proliferation; haematopoiesis; wound healing;  
KW angiogenesis; gene; ds.  
XX Homo sapiens.  
XX OS  
XX WO200272770-A2.  
XX PD 19-SEP-2002.  
XX 08-MAR-2002; 2002WO-US007283.  
XX PF  
XX PR 08-MAR-2001; 2001US-0274281P.  
XX PR 09-MAR-2001; 2001US-0274849P.  
XX PR 12-MAR-2001; 2001US-0275235P.  
XX PR 13-MAR-2001; 2001US-0275579P.  
XX PR 13-MAR-2001; 2001US-0275601P.  
XX PR 14-MAR-2001; 2001US-0276000P.  
XX PR 20-MAR-2001; 2001US-0277239P.  
XX PR 20-MAR-2001; 2001US-0277327P.  
XX PR 20-MAR-2001; 2001US-0277338P.  
XX PR 21-MAR-2001; 2001US-0277791P.  
XX PR 22-MAR-2001; 2001US-0277833P.  
XX PR 23-MAR-2001; 2001US-0278152P.  
XX PR 26-MAR-2001; 2001US-0278894P.  
XX PR 27-MAR-2001; 2001US-0279036P.  
XX PR 28-MAR-2001; 2001US-0279344P.  
XX PR 30-MAR-2001; 2001US-0280233P.  
XX PR 02-APR-2001; 2001US-0280802P.  
XX PR 02-MAY-2001; 2001US-0288148P.  
XX PR 31-MAY-2001; 2001US-0294821P.  
XX PR 31-OCT-2001; 2001US-0335302P.  
XX PR 04-DEC-2001; 2001US-0338375P.  
XX PR 07-MAR-2002; 2002US-00094466.  
XX (CURA-) CURAGEN CORP.  
XX PA  
XX PI Spytek KX, Vernet CA, Tchernev VT, Malyankar UM, Gerlach VL;  
PI Li L, Zerhusen BD, Patturajan M, Gusev VY, Kekuda R, Pena CEA;  
PI Zhong M, Gangoli EA, Taupier RJ;  
XX WPI; 2002-713508/77.  
DR P-PSDB; ABG97480.



XX New NOVX polypeptides and polynucleotides, useful for preventing, PT  
PT diagnosing or treating NOVX-associated disorders, e.g. diabetes, multiple PT  
PT sclerosis, atherosclerosis, cancer, infections, osteoporosis or PT  
XX Parkinson's disease.  
PS Claim 22; Page 98; 266pp; English.  
PS  
CC The present invention relates to a new polypeptide (NOVX). The NOVX CC  
CC polypeptide, nucleic acid and antibody are useful in the manufacture of a CC  
CC medicament for treating a syndrome associated with a human disease, CC  
CC preferably a NOVX-associated disorder. The NOVX nucleic acids, CC  
CC polypeptides and antibodies are useful for treating, preventing or CC  
CC diagnosing diseases such as cancers, Hodgkin disease, Von Hippel-Lindau CC  
CC syndrome, Alzheimer's disease, stroke, tuberculous sclerosis, CC  
CC hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral CC  
CC palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia- CC  
CC telangiectasia, leukodystrophies, addiction, anxiety, depression, pain, CC  
CC obesity, Crohn's disease, osteoporosis, inflammatory bowel disease, CC  
CC scleroderma, haemophilia, diabetes, pancreatitis, autoimmune disease, CC  
CC infections, or graft-versus-host disease. The nucleic acids and CC  
CC polypeptides may also be used as targets for the identification of small CC  
CC molecules that modulate or inhibit e.g. neurogenesis, cell CC  
CC differentiation, cell proliferation, haematopoiesis, wound healing and CC  
CC angiogenesis; in gene therapy, in generation of antibodies that bind CC  
CC immunospecifically to NOVX substances for use in therapeutic or CC  
CC diagnostic methods. The nucleic acids are further used as hybridisation CC  
CC probes, in chromosome mapping, tissue typing, preventive medicine, and CC  
CC pharmacogenomics. The present nucleic acid sequence encodes a human NOVX CC  
XX protein of the invention  
SQ Sequence 1770 BP; 492 A; 373 C; 413 G; 492 T; 0 U; 0 Other;  
Query Match 70.5%; Score 1457.6; DB 6; Length 1770;  
Best Local Similarity 86.2%; Pred. No. 0;  
Matches 1743; Conservative 0; Mismatches 9; Indels 270; Gaps 1;  
QY 47 GCGGTGCGTTAGTTGCGCCCAATGGCGGACCGGTCTCACAGCGCTTTCGCGGAGA 106  
DB 18 GCGGTGCGGTAGTTGCGCCCAATGGCGGACCGGTCTCACAGCGCTTTCGCGGAGA 77  
QY 107 TGGCGCGCTTCGTCCTGTCAGTTAAGACTGGCGGCGGTGAGGACTGGGATTCAAA 166  
DB 78 TGGCGCGCTTCGTCCTGTCAGTTAAGACTGGCGGCGGTGAGGACTGGGATTCAAA 137  
QY 167 TATCGGTGCAATTAGAGATGACTTTTCAATTCTCCCCCAAGAAAACCTGTTGCGTTGG 226  
DB 138 TATCGGTGCAATTAGAGATGACTTTTCAATTCTCCCCCAAGAAAACCTGTTGCGTTGG 197  
QY 227 TGGAACTGTGACAGAGTCTTGTGTAAGTACAAAAGGGTGAACAATGACTTTGAGTT 286  
DB 198 TGGAACTGTGACAGAGTCTTGTGTAAGTACAAAAGGGTGAACAATGACTTTGAGTT 257  
QY 287 GTTGAAGAACCGAGCTGTAGATCAGACATTAAGAGTACCGAGATCAATCGGCTGCT 346  
DB 258 GTTGAAGAACCGAGCTGTAGATCAGACATTAAGAGTACCGAGATCAATCGGCTGCT 317  
QY 347 AGAATTCGGTTCTTCTATCATGTACTTGAACAAAAGACTTTGAGCAACTTATCAGTATTAT 406  
DB 318 AGAATTCGGTTCTTCTATCATGTACTTGAACAAAAGACTTTGAGCAACTTATCAGTATTAT 377  
QY 407 ATTAAGATTGCGTTGGTTGAATAGAGTCAAAACAGTAGTAGTGAAGAGTATTGGCTTTTCT 466  
DB 378 ATTAAGATTGCGTTGGTTGAATAGAGTCAAAACAGTAGTAGTGAAGAGTATTGGCTTTTCT 437  
QY 467 TGGTAATCTGTATCAGCAGACTGTTTTCCTCAGACCGGTGCTCAGCATGATTGCTTC 526  
DB 438 TGGTAATCTGTATCAGCAGACTGTTTTCCTCAGACCGGTGCTCAGCATGATTGCTTC 497  
QY 527 CCAATTTTGTGCTCCCCGAGTGATCAATTAAGGAAGGCGGTAGATGTTTCAGATTCCTGA 586  
DB  
DB 1187 TGGACCAATTTTACTCAGCGCTGCCAAGCTGTGTCTACACCTTTGTTTATAGACACAGCA 1187  
QY 1487 GCTTTTGTAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGCTGTAATTTTGTAGCG 1546  
DB 1128 TGGACCAATTTTACTCAGCGCTGCCAAGCTGTGTCTACACCTTTGTTTATAGACACAGCA 1187  
QY 1487 GCTTTTGTAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGCTGTAATTTTGTAGCG 1546  
DB 1128 TGGACCAATTTTACTCAGCGCTGCCAAGCTGTGTCTACACCTTTGTTTATAGACACAGCA 1187  
QY 1547 GATAGTAGTAGCGCAGCTAAATCCCTGAAAGATTTCCTGCCCTCAGTGTGTTAACTTTT 1606  
DB 1248 GATAGTAGTAGCGCAGCTAAATCCCTGAAAGATTTCCTGCCCTCAGTGTGTTAACTTTT 1307  
QY 1607 TGTGTCAATCAAAATAAGTACAGCTGCTTCTCTCTACACCATCATTTGAGAGGACAA 1666  
DB 1308 TGTGTCAATCAAAATAAGTACAGCTGCTTCTCTCTACACCATCATTTGAGAGGACAA 1367

DB 498 CCATTTTGTGCTCCCGAGTGATCATTAAGCAAGCGATAGATGTTTCAGATTCCTGA 557  
QY 587 TCATGAAGATGATAATCTTCCTGCAAAATTTTGCACATGTCACAGAGCCTTGCAGAAAT 646  
DB 558 TGATGAAGATGATAATCTTCCTGCAAAATTTTGCACATGTCACAGAGCCTTGCAGAAAT 617  
QY 647 AGCAAGATATGTACCATCGACACCGCTGGTTCCTCATGCCAATACCTGCTGGAATAATTC 706  
DB 618 AGCAAGATATGTACCATCGACACCGCTGGTTCCTCATGCCAATACCTGCTGGAATAATTC 677  
QY 707 ATTGTTTGAATAATCAGAGAGAACCTGGAATGTTTACGTTTCACTTACTAAGGATTAG 766  
DB 678 ATTGTTTGAATAATCAGAGAGAACCTGGAATGTTTACGTTTCACTTACTAAGGATTAG 737  
QY 767 TGATATTTTCAACCTTGAGGCATGAATCTGAGCTTATATGAAAACTACTCAA 826  
DB 738 TGATATTTTCAACCTTGAGGCATGAATCTGAGCTTATATGAAAACTACTCAA 797  
QY 827 GTTGATGTGAATGCCGCGAGGTATTAAGATGCTGAAGAAACAGCAACTCAAC 886  
DB 798 GCTGATGTGAATGCATCCCGCAGGTTTGAAGATGCTGAAGAAACAGCAACTCAAC 857  
QY 887 TTGCTGTGGACAGATTCCACGGAAGGATTGTTAATATGATGAAGATGAAGAACTCA 946  
DB 858 TTGCTGTGGACAGATTCCACGGAAGGATTGTTAATATGATGAAGATGAAGAACTCA 895  
QY 947 ACATGAAACAAAGGCTGCTGACCGCTCGACAGATGTTGATCTCTGTAGCGAGG 1006  
DB 896 ----- 895  
QY 1007 CTTGACATCTGTATGTTTGGTTCCTTCTCATCATGAAGGATGCTGCTATGTAGTGG 1066  
DB 896 ----- 895  
QY 1067 TAAGTTGATTAACGGCAAAAACAAAGGATCTATATGCGACCTGATAAACAATTTTGCAA 1126  
DB 896 ----- 895  
QY 1127 ACTCCTGTTGCCACCCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1186  
DB 896 ----- 895  
QY 1187 TTTCAAAATTTGGATTCGACAGGCAATTTTGGAAACATCTCTGGAATAAATTTGACAGCC 1246  
DB 896 -----TGCGATTCGACAGGCAATTTTGGAAACATCTTGGAAATAAATTTGACAGATCC 947  
QY 1247 AAGTAATCTGCTCATCATGACGCAAGGCTGCTGGAATAATATATTTGGAAGCTTTTGGCAAG 1306  
DB 948 AAGTAATCTGCTCATCATGACGCAAGGCTGCTGGAATAATATATTTGGAAGCTTTTGGCAAG 1007  
QY 1307 AGCTAAATTTTATCTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1366  
DB 1008 AGCTAAATTTTATTTCTTTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1067  
QY 1367 GCACATATACCTTAAATAACCAAGGATTCGGGAAACAAAGGCAATCTGCGATGTTGCTCTCCA 1426  
DB 1068 GCACATATACCTTAAATAACCAAGGATTCGGGAAACAAAGGCAATCTGCGATGTTGCTCTCCA 1127  
QY 1427 TGGACCAATTTTACTCAGCGCTGCCAAGCTGTGTCTACACCTTTGTTTATAGACACAGCA 1486  
DB 1128 TGGACCAATTTTACTCAGCGCTGCCAAGCTGTGTCTACACCTTTGTTTATAGACACAGCA 1187  
QY 1487 GCTTTTGTAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGCTGTAATTTTGTAGCG 1546  
DB 1128 GCTTTTGTAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGCTGTAATTTTGTAGCG 1187  
QY 1547 GATAGTAGTAGCGCAGCTAAATCCCTGAAAGATTTCCTGCCCTCAGTGTGTTAACTTTT 1606  
DB 1248 GATAGTAGTAGCGCAGCTAAATCCCTGAAAGATTTCCTGCCCTCAGTGTGTTAACTTTT 1307  
QY 1607 TGTGTCAATCAAAATAAGTACAGCTGCTTCTCTCTACACCATCATTTGAGAGGACAA 1666  
DB 1308 TGTGTCAATCAAAATAAGTACAGCTGCTTCTCTCTACACCATCATTTGAGAGGACAA 1367

1667 TCGCAGATGTCGACGATCATTTAGAGTACCGCTGGAGAGACTCAGTGCAGATCTGCAC 1726  
1368 TCGCAGATGTCGACGATCATTTAGAGTACCGCTGGAGAGACTCAGTGCAGATCTGCAC 1427  
1727 AAACCCGCTGGACACCTTCTCCCTTTGATCCCTGTGCTGAGAGAGTCAAAGAAATT 1786  
1428 AAACCCGCTGGACACCTTCTCCCTTTGATCCCTGTGCTGAGAGAGTCAAAGAAATT 1487  
1787 CATTGATCCTATTTATCAGGTATGGAGACATGATGCTGAGAGCTACAGAGTTCAA 1846  
1498 CATTGATCCTATTTATCAGGTATGGAGACATGATGCTGAGAGCTACAGAGTTCAA 1547  
1847 GAAACCCATGAAAGGACATAGTGGAGATCAAGATGATGATCTTCTGAAAGGCGAAGT 1906  
1548 GAAACCCATGAAAGGACATAGTGGAGATCAAGATGATGATCTTCTGAAAGGCGAAGT 1607  
1907 GCCCCAGATGATACCGTATGGGATCACACCAAGCTCCTTTGACACGCAATTTCCGAAG 1966  
1608 GCCCCAGATGATACCGTATGGGATCACACCAAGCTCCTTTGACACGCAATTTCCGAAG 1667  
1967 TCCTTCAAGTACTGTGGCTCCCAACCCGCTGTTGTATGACATGCAACCCAGTCCCTCTGAGC 2026  
1668 TCCTTCAAGTACTGTGGCTCCCAACCCGCTGTTGTATGACATGCAACCCAGTCCCTCTGAGC 1727  
2027 GCAGAAATTTGACTGAGATGACATTTGGGATTTCCCAT 2068  
1728 GCAGAAATTTGACTGAGATGACATTTGGGATTTCCCAT 1769

RESULT 4  
ADD29697  
ID ADD29697 standard; mRNA; 1582 BP.  
XX  
AC ADD29697;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Human tumour suppressor mRNA SEQ ID NO:153.  
XX  
KW ss; human; tumour suppressor; cancer; cancer; cytostatic; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO2003058201-A2.  
XX  
PD 17-JUL-2003.  
XX  
PF 31-DEC-2002; 2002WO-US041825.  
XX  
PX 31-DEC-2001; 2001US-0345317P.  
XX  
PY {QUAR-} QUARK BIOTECH INC.  
XX  
PA {CLEV-} CLEVELAND CLINIC FOUND.  
XX  
PI Feinstein E, Gudkov AV;  
XX  
PW WPI; 2003-598393/56.  
XX  
PT Diagnosing cancer comprises determining the polypeptide or polynucleotide  
PT levels e.g., hepatic lipase, in a sample from a subject, where a higher  
PT level compared to that in a subject free of cancer is indicative of  
PT cancer.  
XX  
PS Disclosure; SEQ ID NO 153; 272pp; English.  
XX  
XX The invention relates to a novel method for diagnosing a cancer in a  
XX subject, the method comprises determining, in a sample from the subject,  
XX the level of at least one polypeptide, where a higher level of the  
XX polypeptide compared to the level of the polypeptide in a subject free of  
XX cancer is indicative of cancer. The polypeptide is selected from any of  
XX the polypeptides encoded by the polynucleotides listed in the  
XX specification and polypeptides which are at least 70% homologous to the

CC polypeptides. The method of the invention has cytostatic activity, and  
CC may have a use in gene therapy. The method is useful in identifying  
CC markers specific for one or several types of cancer, depending on the  
CC tissue origin, which may be used in numerous diagnostic and prognostic  
CC applications as well as cancer type-specific targets for therapeutic  
CC intervention. The compounds that modulate the activity of a tumour  
CC suppressor gene are useful in the treatment of cancer or as anti-cancer  
CC drugs. The present sequence represents a polynucleotide of the invention.  
XX  
SQ Sequence 1582 BP; 364 A; 253 C; 296 G; 389 T; 0 U; 280 Other;  
Query Match 60.6%; Score 1253.6; DB 9; Length 1582;  
Best Local Similarity 81.0%; Pred. No. 0;  
Matches 1276; Conservative 0; Mismatches 299; Indels 1; Gaps 1;  
QY 47 GCGGTCGGTGTAGTTCCGCCCAATGCGCGCACCGCTGCTTCCACAG-CGTTTTGCCGGGAG 105  
DB 1 GCGGTCGGGTTAGTTTCGGGCCCAATGCGCGCACCGCTGCTTCCACAGTTGTTTGTCCGGAG 60  
QY 106 ATGCGGCGCTTCGCTCTCTGCAGTTAAGAACTGGGCGCTGCGAGGACTGGGATTTCAA 165  
DB 51 ATGCGGCGCTTCGCTCTCTGCAGTTCAAGACGCTGGCGCTGCGAGGACTGGGATTTCAA 120  
QY 166 ATATCGGTGCATTAGAGAAGTATTTTCAATTTCTCCCAAGAAACTGTTTCGTTTG 225  
DB 121 ATATCGGTGCATTAGAGAAGTATTTTCAATTTCTCCCAAGAAACTGTTTCGTTTG 180  
QY 226 GTGGAACTGTGACAGAGTCTTCTGCTGAAAGTACAAAAGGCTGAAACAAATGATTTGAGT 285  
DB 181 GTGGAACTGTGACAGAGTCTTCTGCTGAAAGTACAAAAGGCTGAAACAAATGATTTGAGT 240  
QY 286 TGTTCGAAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACAGATCATCAACTGGCTGC 345  
DB 241 TGTTCGAAAGAACCCAGCTGTTAGATCCAGACATAAAGGATGACAGATCATCAACTGGCTGC 300  
QY 346 TAGAATTCGGTTCCTTCTATCATGTTACTTGCAGAAAGACTTTGAGCAACTTATCAGTATTA 405  
DB 301 TAGAATTCGGTTCCTTCTATCATGTTACTTGCAGAAAGACTTTGAGCAACTTATCAGTATTA 360  
QY 406 TATTAAGATTGCTGCTGTTGAATAGAAAGTCAACACAGTCTAGTGAAGAGTATTTGGCTTTTC 465  
DB 361 TATTAAGATTGCTGCTGTTGAATAGAAAGTCAACACAGTCTAGTGAAGAGTATTTGGCTTTTC 420  
QY 466 TTGGTAAATCTTGTTATCAGCAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTCGTT 525  
DB 421 TTGGTAAATCTTGTTATCAGCAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTCGTT 480  
QY 526 CCAATTTTGTGCTCCCGAGTGATCATTTAAGAAAGCGCATGTAGATGTTTCAGATTCG 585  
DB 481 CCAATTTTGTGCTCCCGAGTGATCATTTAAGAAAGCGCATGTAGATGTTTCAGATTCG 540  
QY 586 ATGATGAAGATGATAATCTTCTGCAAAATTTTGCACATGTCTACAGAGCTTCGCAATAA 645  
DB 541 ATGATGAAGATGATAATCTTCTGCAAAATTTTGCACATGTCTACAGAGCTTCGCAATAA 600  
QY 646 TAGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATFACCTGGTGGAAAAATTC 705  
DB 601 TAGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATFACCTGGTGGAAAAATTC 660  
QY 706 CATTGTTTGAATAATCAGAGAGAACTGGAATGTTTACGTTTCACTTACTTAAGGATTA 765  
DB 661 CATTGTTTGAATAATCAGAGAGAACTGGAATGTTTACGTTTCACTTACTTAAGGATTA 720  
QY 766 GTGTATATTTTCCAACTTGAGGCATCAAAATCTGGAGCTTATTTATGAAAACTACTCA 825  
DB 721 GTGTATATTTTCCAACTTGAGGCATCAAAATCTGGAGCTTATTTATGAAAACTACTCA 780  
QY 826 AGTTGGATGTGAATGATCCCGCAGGGTATTGAAGATGCTGAAGAAACGCAACTCAA 885  
DB 781 AGCTGGATGTGAATGATCCCGCAGGGTATTGAAGATGCTGAAGAAACGCAACTCAA 840  
QY 886 CTTTGGTGGGACAGATTCACCGAGGATGTTTATATATGATGAAGATGAAGAAACTG 945

Db 841 CTTGTGTTGGGACAGATTCCAGGAAGGATGTTTAAATGNNNNNNNNNNNNNNNNNN 900  
QY 946 AACATGAAACAAAGGCTGTCCTGAACGGCTGACGAGATGGTGCATCTGTAGCCGAGC 1005  
Db 901 NNN 960  
QY 1006 GCGTGACATCCTGATGTCCTTGTTGTTGTCCTACATGAAGGATGTCGTATGTAGATG 1065  
Db 961 NNN 1020  
QY 1066 GTAAGGTTGATAACCGGCAAAACAAAGATCTATATCGGACCTGATAAACATCTTTGACA 1125  
Db 1021 NNN 1080  
QY 1126 AACTCCTGTTGCCACCATGCTCTCGCATGTACAGTTTTTCATGTTTTACCTCTGTA 1185  
Db 1081 NNN 1140  
QY 1186 GTTTCAAATGGGATTGCGACAGGCAATTTTGGAAACATCTCTGGAATAATTCGAGGACC 1245  
Db 1141 NNN 1200  
QY 1246 CAAAGTAATCCTCCATCATCAGCAGGCTGCTGGAATATATATGGAAGCTTTTGGCAA 1305  
Db 1201 CAAAGTAATCCTCCATCATCAGCAGGCTGCTGGAATATATATGGAAGCTTTTGGCAA 1260  
QY 1306 GAGCTAAATTTATCTCTTATCTATCTGTAATCATGCTAGATCTTTGGTTAACTGGC 1365  
Db 1261 GAGCTAAATTTATCTCTTATCTATCTGTAATCATGCTAGATCTTTGGTTAACTGGC 1320  
QY 1366 TGCACATATACCTTAATACACGAGATTCGGGAACAAAGGCATCTCGCATGTTGCTCTCC 1425  
Db 1321 TGCACATATACCTTAATACACGAGATTCGGGAACAAAGGCATCTCGCATGTTGCTCTCC 1380  
QY 1426 ATGACCAATTTACTCAGCTGCCAAGCTGTTCTTACACCTTGTCTTACACCAAGC 1485  
Db 1381 ATGACCAATTTACTCAGCTGCCAAGCTGTTCTTACACCTTGTCTTACACCAAGC 1440  
QY 1486 AGCTTTTACGCGGAACCTGAAAGAGTTTGCAGTATCTTACAGTCTGAAATTTGAGC 1545  
Db 1441 AGCTTTTACGCGGAACCTGAAAGAGTTTGCAGTATCTTACAGTCTGAAATTTGAGC 1500  
QY 1546 GGATAGTATGATGACGAGCAATATCCCTGAAGATTTGCTGCTCCCTCAGTGTAACTTTT 1605  
Db 1501 GGATAGTATGATGACGAGCAATATCCCTGAAGATTTGCTGCTCCCTCAGTGTAACTTTT 1560  
QY 1606 TTGCTGCAATCACAAA 1621  
Db 1561 TTGCTGCAATCACAAA 1576

## RESULT 5

AACT6937 standard; cDNA; 1418 BP.

AACT6937;

08-FEB-2001 (first entry)

Human ORFX ORF2492 polynucleotide sequence SEQ ID NO:4983.

Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;  
vulnerable; antipsoriatic; antiparkinsonian; nontropic; neuroprotective;  
anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiac;  
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
hypotensive; dermatological; immunosuppressive; antiinflammatory;  
antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
neurodegenerative disorder; osteoarthritis; graft vs host disease;  
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
cholesterol ester storage; systemic lupus erythematosus; infection;  
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

bone damage; cartilage damage; antiinflammatory disease; coagulation;  
thrombosis; contraceptive; ss.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US008621.

31-MAR-1999; 99US-0127607P.

02-APR-1999; 99US-0127636P.

05-APR-1999; 99US-0127728P.

30-MAR-2000; 2000US-00540763.

(CURA-) CURAGEN CORP.

Shinkets RA, Leach M;

WPI; 2000-602362/57.

P-PSDB; AAB42728.

Novel nucleic acids and peptides derived from open reading frame X,

useful for treating e.g. cancers, proliferative disorders,

neurodegenerative disorders and cardiovascular disease.

Claim 5; Page 4157-4158; 5507pp; English.

AACT6937 encode the proteins given in AAB40237 to AAB43397,  
which represent the human ORFX open reading frames 1 to 3161. The ORFX  
sequences have activities such as: cytostatic; hepatotropic; vulnerable;  
antipsoriatic; antiparkinsonian; nontropic; neuroprotective; osteopathic;  
anticonvulsant; antirheumatic; immunosuppressant; antidiabetic; hypotensive;  
cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; antithyroid;  
antianaemic; antirheumatic; antirheumatic; antithyroid; antianaemic. The  
sequences can be used for determining the presence of or predisposition  
to, or preventing or treating pathological conditions associated with an  
ORFX-associated disorder. The nucleic acids can be used to express ORFX  
proteins in gene therapy vectors. The proteins and nucleic acids may be  
used to treat cancers, proliferative disorders, neurodegenerative  
disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
storage, systemic lupus erythematosus, severe combined immunodeficiency  
(SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
enhance coagulation; to inhibit thrombosis; and as a contraceptive

Sequence 1418 BP; 422 A; 275 C; 318 G; 403 T; 0 U; 0 Other;

Query Match 59.4%; Score 1228.4; DB 3; Length 1418;

Best Local Similarity 98.7%; Pred. No. 0;

Matches 1249; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 670 CTTGTTTCTCATGCCCAATCTGTCGAAATTTCCATTTGTCGAAATTCAGAGAA 729

Db 1 CTTGTTTCTCATGCCCAATCTGTCGAAATTTCCATTTGTCGAAATTCAGAGAA 60

QY 730 CACTGGAATGTTACGTTTCTAATCTAAGGATAGTGTATATTTTCCAA-CCTTCAGG 788

Db 61 CACTGGAATGTTACGTTTCTAATCTAAGGATAGTGTATATTTTCCAACTTCAGG 120

QY 789 CATGAATTCGAGCTTATTTTGAATACTACTCAAGTTGGATGTAATGCATCCCG 848

Db 121 CATGAATTCGAGCTTATTTTGAATACTACTCAAGTTGGATGTAATGCATCCCG 180

QY 849 CAGGTTTCAAGATGCTGAGAGAACAGCACTCAAACTTGTGTGGGACAGATTCACG 908

Db 181 CAGGTTTCAAGATGCTGAGAGAACAGCACTCAAACTTGTGTGGGACAGATTCACG 240

QY 909 GAAGGATGTTTAAATATGATGAAGATGAAGAACTGAACATCAACAAAGGCTGCTCT 968



Query Match 48.9%; Score 1010.4; DB 4; Length 1498;  
Best Local Similarity 98.8%; Pred. No. 2.8e-289;  
Matches 1028; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

```
QY 735 GAAATGTTAGCTTCATAACTACTAAGGATTAGTGATATATTTTCCAACTTGAGGCGATGAA 794
DB 4 GAAAGTTACGTTTCATAACTTACAGAGATTAGTGATATATTTTCCAACTTGAGGCGATGAA 62
QY 795 ATTCTGGAGCTTATTTTGAAGAACTACTCAAGTTGGATGTGAATGCATCCCGGACGGGT 854
DB 63 ATTCTGAGCTTATTTATGAGAACTACTCAAGTTGGATGTGAATGCATCCCGGACGGGT 122
QY 855 ATTGAAGATGCTGAAGAAACAGCACTCAAACTTGGTGGGACAGATTCCACGGAAGA 914
DB 123 ATTGAAGATGCTGAAGAAACAGCACTCAAACTTGGTGGGACAGATTCCACGGAAGA 182
QY 915 TTGTTTAAATATGATGAAGATGAAGAACTGAACATGAACAAAGGCTGGTCTGAACGG 974
DB 183 TTGTTTAAATATGATGAAGATGAAGAACTGAACATGAACAAAGGCTGGTCTGAACGG 242
QY 975 CTCGACAGATGGTGATCTCTAGCCGAGCCCTGGACATCTCTGATGTCTTTGGTTTGG 1034
DB 243 CTCGACAGATGGTGATCTCTAGCCGAGCCCTGGACATCTCTGATGTCTTTGGTTTGG 302
QY 1035 TCCTACATGAAGGATGCTCTGCTATGTAGATGTAGTGAAGTTGATAAGCGCAAAACAAAGGAT 1094
DB 303 TCCTACATGAAGGATGCTCTGCTATGTAGATGTAGTGAAGTTGATAAGCGCAAAACAAAGGAT 362
QY 1095 CTATATCGCGACTGATAAACATCTTTGACAACTCTCTGTTGCCACCCATCGCTCCCTGC 1154
DB 363 CTATATCGCGACTGATAAACATCTTTGACAACTCTCTGTTGCCACCCATCGCTCCCTGC 422
QY 1155 CATGTACAGTTTTCATGTTTACCTCTGTAGTTTCAAAATGGGATTCGACAGAGCATTT 1214
DB 423 CATGTACAGTTTTCATGTTTACCTCTGTAGTTTCAAAATGGGATTCGACAGAGCATTT 482
QY 1215 TTGGAACATCTCTGAAAGAAATTCGAGACCCCAAGTAATCTTGCCATCATCAGGAGGCT 1274
DB 483 TTGGAACATCTCTGAAAGAAATTCGAGACCCCAAGTAATCTTGCCATCATCAGGAGGCT 542
QY 1275 GCTGAAATATATTTGGAAGCTTTTGGCAAGAGCTAAATTTATTCCTCTTATTAAGTGA 1334
DB 543 GCTGAAATATATTTGGAAGCTTTTGGCAAGAGCTAAATTTATTCCTCTTATTAAGTGA 602
QY 1335 AAATCATGCTAGATCTTTTGGTTAACTGGTGACATATATACCTTAATAACAGGATTCG 1394
DB 603 AAATCATGCTAGATCTTTTGGTTAACTGGTGACATATATACCTTAATAACAGGATTCG 662
QY 1395 GGAACAAAGGATCTGCGATGTTGCTCTCCATGACCACTTTTACTGAGCTGCGAAGCT 1454
DB 663 GGAACAAAGGATCTGCGATGTTGCTCTCCATGACCACTTTTACTGAGCTGCGAAGCT 722
QY 1455 GTGTTCTACACTTTGTTTATGACAAAGCAGCTTTTGGCGGAAACCTGAAAGAGGT 1514
DB 723 GTGTTCTACACTTTGTTTATGACAAAGCAGCTTTTGGCGGAAACCTGAAAGAGGT 782
QY 1515 TTGACAGTATCTCAGAGTCTGAAATTTGAGCGGATAGTGATGAGCCAGCTAAATCCCTCG 1574
DB 783 TTGACAGTATCTCAGAGTCTGAAATTTGAGCGGATAGTGATGAGCCAGCTAAATCCCTCG 842
QY 1575 AAGATTTGCTGCTCAGTGGTTAACTTTTTCGTCGAATCAAAATAGTACAGCTC 1634
DB 843 AAGATTTGCTGCTCAGTGGTTAACTTTTTCGTCGAATCAAAATAGTACAGCTC 902
QY 1635 GTCTTCTGTACACATCATGAGAGGAAACAACTGCGAGATGCTGCGAGCTATTAGAGT 1694
DB 903 GTCTTCTGTACACATCATGAGAGGAAACAACTGCGAGATGCTGCGAGCTATTAGAGT 962
QY 1695 ACCCTGAGAGAGCTCAGTGCAGATCTGCAAAACCCGCTGGACACTTCTTCCCTCTTT 1754
DB 963 ACCCTGAGAGAGCTCAGTGCAGATCTGCAAAACCCGCTGGACACTTCTTCCCTCTTT 1022
```

```
QY 1755 GATCCCTGTGTGCTGAAGAG 1774
DB 1023 GATCCCTGTGTGCTGAAGAG 1042
```

```
RESULT 7
AAS01562
ID AAS01562 standard; cDNA; 1423 BP.
XX AAS01562;
XX AC
XX 18-JUL-2001 (first entry)
XX DE
XX Human secretory molecule cDNA sptm #52.
```

Human; secretory molecule; sptm; SPTM; library screening; gene therapy;  
cell signalling; cell proliferative disorder; atherosclerosis; cancer;  
immune system disorder; AIDS; neurological disorder; Alzheimer's disease;  
nervous system disease; mental retardation; developmental disorder;  
neuromuscular disorder; microarray; Incyte ID number 4287452dec; ss.

OS Homo sapiens.

XX WO200123558-A2.

XX 05-APR-2001.

XX 19-SEP-2000; 2000WO-US025610.

XX 28-SEP-1999; 99US-0156624P.

XX 28-SEP-1999; 99US-0156625P.

XX 02-DEC-1999; 99US-0168611P.

XX 02-DEC-1999; 99US-0168613P.

XX 02-DEC-1999; 99US-0168614P.

XX (INCY-) INCYTE GENOMICS INC.

XX Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banville SC;

XX Bratcher SR, Dufour GE, Cohen HJ, Rosen BH, Shah P, Chalup MS;

XX Hillman JL, Jones AL, Yu JY, Greenawalt LB, Panzer SR, Roseberry AM;

XX Wright RJ, Chen W, Liu TP, Yap PE, Stockdreher TK, Amshey S;

XX Fong WT;

XX WPI; 2001-258134/26.

XX New secretory polynucleotides (SPTM) and the polypeptides they encode,

XX for use in inducing antibodies and screening libraries of compounds.

XX Claim 1; Page 152; 161pp; English.

The present sequence for human secretory molecule cDNA sptm #52 (Incyte ID number 4287452dec) is 1 of 63 novel sptm cDNA sequences (AAS01511-AAS01573) which encode for the secretory polypeptides SPTM. The sptm polynucleotides are useful for screening a compound for effectiveness in altering expression of a target polynucleotide, where the target polynucleotide comprises sptm. Sptm is also useful in a method for assessing the toxicity of a test compound. Sptm and its fragments or complementary sequences, may be used to identify the presence of and/or determine the degree of similarity between two nucleic acid sequences. Sptm can also be used for a variety of diagnostic and therapeutic purposes, e.g. diagnosing a particular condition, disease or disorder associated with cell signalling, such conditions include cell proliferative disorders such as atherosclerosis, and cancers including leukaemia, an immune system disorder e.g. acquired immunodeficiency syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's disease, nutritional and metabolic disease of the nervous system, mental retardation and other developmental disorders, and muscular dystrophy and other neuromuscular disorders. Sptm can also be used to design probes useful in diagnostic assays, which may be used to monitor the progress of conditions or disorders associated with abnormal levels of expression of sptm. In addition sptm encoding SPTM may be used for somatic or germline gene therapy, for inducing antibodies, or in microarrays

XX

SQ	Sequence 1423 BP; 417 A; 247 C; 320 G; 439 T; 0 U; 0 Other;	
	Query Match 43.2%; Score 894; DB 4; Length 1423;	
	Best Local Similarity 80.9%; Pred. No. 1.2e-254;	
	Matches 1191; Conservative 0; Mismatches 10; Indels 271; Gaps 2;	
QY	157 GGAATTCCTCAATATGCGTATAGAGATGACCTTTTCAATTTCTCCCCACAGAAAACATG	216
DB	192 GGATTTCAAAATATGGGTGCAATAGAGATGATTTTTCATTTCTCCCCACAGAAAACATG	251
QY	217 TTCCGTTTGGTGGAACTGTGACAGAGTCTTGCTGAAGTACAAAAGGGGTGAAAACATG	276
DB	252 TTCCGTTTGGTGGAACTGTGACAGAGTCTTGCTGAAGTACAAAAGGGGTGAAAACATG	311
QY	277 ACTTTGAGTTTGAAGAACCGAGCTGTAGATCCAGACATTAAGATGACCGAGATCATCA	336
DB	312 ACTTTGAGTTTGAAGAACCGAGCTGTAGATCCAGACATTAAGATGACCGAGATCATCA	371
QY	337 ACTGGCTGCTAGAATTCGGTCTCTCTATCATGTACTTGAACAAAGACTTTGAGCACTTA	396
DB	372 ACTGGCTGCTAGAATTCGGTCTCTCTATCATGTACTTGAACAAAGACTTTGAGCACTTA	431
QY	397 TCAGTATATATTAAGATTCGCTTGGTTGAATAGAGTCAAAACAGTAGTGAAGATATT	456
DB	432 TCAGTATATATTAAGATTCGCTTGGTTGAATAGAGTCAAAACAGTAGTGAAGATATT	491
QY	457 TGGCTTTCTTGGTAACTGTATCAGACAGACTGTTTTCCTCAGACCGTGTCTCAGCA	516
DB	492 TGGCTTTCTTGGTAACTGTATCAGACAGACTGTTTTCCTCAGACCGTGTCTCAGCA	551
QY	517 TGATTCCTTCCCATTTTGGCTCCCGAGTGATCATTAAGGAAGCGATGTAGATGTTT	576
DB	552 TGATTCCTTCCCATTTTGGCTCCCGAGTGATCATTAAGGAAGCGATGTAGATGTTT	611
QY	577 CAGATTCCTGATGATGAAGATGATAATCTTTCCTGCAAAATTTTGACATGTCACAGAGCCT	636
DB	612 CAGATTCCTGATGATGAAGATGATAATCTTTCCTGCAAAATTTTGACATGTCACAGAGCCT	671
QY	637 TGCATAATAGCAAGATATGACATCGACACCGTGGTTTCTCATGCCAATACTCGTGG	696
DB	672 TGCATAATAGCAAGATATGACATCGACACCGTGGTTTCTCATGCCAATACTCGTGG	731
QY	697 AAAAAATTCATTTGTTCCGAAATCAGAGAAACACTGGAAATGTACGTTTCAATACCTTAC	756
DB	732 AAAAAATTCATTTGTTCCGAAATCAGAGAAACACTGGAAATGTACGTTTCAATACCTTAC	791
QY	757 TAAGGATTAAGTATATTTTCCAACTTGAGGCATGAAATTCGGAGCTTTATTGAAA	816
DB	792 TAAGGATTAAGTATATTTTCCAACTTGAGGCATGAAATTCGGAGCTTTATTGAAA	851
QY	817 AACTACTCAAGTTGGATGTAATGCAATCCCGCAGGGTATTGAAGATGCTGAAGAAACAG	876
DB	852 AACTACTCAAGTTGGATGTAATGCAATCCCGCAGGGTATTGAAGATGCTGAAGAAACAG	911
QY	877 CAATCTCAAACTTGGTGGACAGATTCACGGAAGGATTTGTTTAAATATGATGAAGATG	936
DB	912 CAATCTCAAACTTGGTGGACAGATTCACGGAAGGATTTGTTTAAATATGATGAAGATG	959
QY	937 AAGAACTGAACATGAACAAAGAGCTGGTCTCTGAAACCGCTCGACCATGGTGCATCCTG	996
DB	960 -----	959
QY	997 TAGCCGAGCGCTGGACATCCTGATGCTTTGGTTTGTCTCTCATGAAGGATGCTGCT	1056
DB	960 -----	959
QY	1057 ATGTAGATGTTAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACA	1116
DB	960 -----	959
QY	1117 TCCTTGACAAACTCCTGTTGCCACCCATGCCTCCTGCCATGTACAGTTTTCATGTTTT	1176
DB	960 -----	959

QY	1177	ACCTCTGATGTTTCAAAATTTGGATTTCGCAGAGG-CATTTTGTGAAACATCTCTGGAATAAA	1235
DB	960	-----TGGATTTCGCAGAGGCCATTTTGGAAACATCTTTTGGAAATAAC	1001
QY	1236	TTGCAGGACCCCAAGTAATCTCTGCCATCATCAGGAGGCTGCTGGAAATATATTTGGAAGC	1295
DB	1002	TTGCAGGATCCCAAGTAATCTCTGCCATCATCAGGAGGCTGCTGGAAATATATTTGGAAGC	1061
QY	1296	TTTTTGGCAAGAGCTAATAATTTATTTCTCTTACTGTAAATCATGCTAGATCTTTTG	1355
DB	1062	TTTTTGGCAAGAGCTAATAATTTATTTCTTACTGTAAATCATGCTAGATCTTTTG	1121
QY	1356	GTTAACTGGCTGCACATATACCTTAATAACACAGGATTCGGGAACAAAGGCATTTCTGCGAT	1415
DB	1122	GTTAACTGGCTGCACATATACCTTAATAACACAGGATTCGGGAACAAAGGCATTTCTGCGAT	1181
QY	1416	GTTGCTCTCCATGGACCAATTTTACTCAGCTCCGAAAGTGTGTTCTACACCTTTGTTTT	1475
DB	1182	GTTGCTCTCCATGGACCAATTTTACTCAGCTCCGAAAGTGTGTTCTACACCTTTGTTTT	1241
QY	1476	AGACACAGCAGCTTTTGAGCGGAACCTGAAAGAGGTTTCAGTATCTTCAGAGTCTG	1535
DB	1242	AGACACAGCAGCTTTTGAGCGGAACCTGAAAGAGGTTTCAGTATCTTCAGAGTCTG	1301
QY	1536	AATTTTGAGCGGATAGTATGAGCCAGCTAAATCCCTGAAAGATTTGCTGCCCTCAGTG	1595
DB	1302	AATTTTGAGCGGATAGTATGAGCCAGCTAAATCCCTGAAAGATTTGCTGCCCTCAGTG	1361
QY	1596	GTTAACTTTTGTGCAATCAAAATAGTA	1627
DB	1362	GTTAACTTTTGTGCAATCAAAAGATGAA	1393

RESULT 8  
AAS92255  
ID AAS92255 standard; cDNA; 2410 BP.  
XX  
AC AAS92255;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #28059.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
XX  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG28068.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 28059; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)

sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 2410 BP; 672 A; 437 C; 547 G; 754 T; 0 U; 0 Other;

Query Match  
Best Local Similarity 30.9%; Score 639.4; DB 5; Length 2410;  
Matches 706; Conservative 0; Mismatches 16; Indels 7; Gaps 5;

QY 1347 GATCTTTGGTTAACTGGCTGCACATATACCTTAATAACAGGATTCGGGAACAAGGCA 1406  
DB 1 GATCTTTGGTTAACTGGCTGCACATATACCTTAATAACAGGATTCGGGAACAAGGCA 60

QY 1407 TTCTCGGATGTGCTCTCGATGGACATTTTACTCAGCTGCCAAGCTGTGTTCACACC 1466  
DB 61 TTCTCGGATGTGCTCTCGATGGACATTTTACTCAGCTGCCAAGCTGTGTTCACACC 120

QY 1467 TTTGTTTTAGACACAAGCAGCTTTTGACGGAAACCTTGAAAGAGCTTGCAGTATCTT 1526  
DB 121 TTTGTTTTAGACACAAGCAGCTTTTGACGGAAACCTTGAAAGAGCTTGCAGTATCTT 180

QY 1527 CAGAGTCTGAATTTTGACGGATAGTATGATGACCAAGCTAAATCCCTCAAGAT-TTGCT 1585  
DB 181 CAGAGTCTGAATTTTGACGGATAGTATGATGACCAAGCTAAATCCCTCAAGATATGGCTT 240

QY 1586 GCCCTCAGTGTAACTTTTGTGCAATCACAAATAGTACAGCTCGTCTTCTGCTA 1645  
DB 241 GCCCTCAGTGTAACTTTTGTGCAATCACAAATAGTACAGCTCGTCTTCTGCTA 300

QY 1646 CACCATCATTCAGAGGAACAATCGCAGATGTCGAGTATTTAGGAGTACCGTGGAGG 1705  
DB 301 CACCATCATTCAGAGGAACAATCGCAGATGTCGAGTATTTAGGAGTACCGTGGAGG 360

QY 1706 AGACTCAGTGCAGATCTGCAAAACCCGCTGGACACCTTCTCCCTTT-GATCCCTGTG 1764  
DB 361 AGACTCAGTGCAGATCTGCAAAACCCGCTGGACACCTTCTCCCTTTGGATCCCTGTG 420

QY 1765 TGCTGAAGAGTCAAGAAAT-CATTGATCCTATT---ATCAGGTATGGGAACAATG 1820  
DB 421 TGCTGAAGAGTCAAGAAATCATTGATCCTATTATCCAGTATGGGGGAACAATG 480

QY 1821 AGTGTCTGAAGAGTACAGAGTTCAGAAACCCATGAAAGGACATAGTGGAGATGAA 1880  
DB 481 AGTGTCTGAAGAGTACAGAGTTCAGAAACCCATGAAAGGACATAGTGGAGATGAA 540

QY 1881 GATGATGATCTTCTGAAAGGCGAAGTGCCTCCAGAAATGATACCGTGAATGGATCACACCA 1940  
DB 541 GATGATGATCTTCTGAAAGGCGAAGTGCCTCCAGAAATGATACCGTGAATGGATCACACCA 600

QY 1941 AGCTCTTTTGACAGCATTTCCGAGAGTCTTCAAGTAGTGTGGGTCCCAACCGTGTG 2000  
DB 601 AGCTCTTTTGACAGCATTTCCGAGAGTCTTCAAGTAGTGTGGGTCCCAACCGTGTG 660

QY 2001 TACA-TGCAACCCAGTCCCTCTCAGCGCAGAAATTTGTGACTGAGATGACATTTGGG 2059  
DB |||||

DB 661 TACACTGCAACCCAGTCCCTCTGACGGCAGAAATTTGCGACTGAGATGTCATTTGG 720  
QY 2060 ATTCCCAT 2068  
DB 721 ATTCCCAT 729

RESULT 9  
AAS92253  
ID AAS92253 standard; cDNA; 776 BP.  
XX AAS92253;  
XX 13-FEB-2002 (first entry)  
DE DNA encoding novel human diagnostic protein #28057.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US008631.  
XX 31-MAR-2000; 2000US-00540217.  
XX 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX P-PSDB; ABG28066.  
XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.  
XX Claim 1; SEQ ID NO 28057; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 776 BP; 206 A; 157 C; 198 G; 215 T; 0 U; 0 Other;

Query Match 27.5%; Score 568.6; DB 5; Length 776;  
Best Local Similarity 97.8%; Pred. No. 5.2e-158;  
Matches 629; Conservative 0; Mismatches 9; Indels 5; Gaps 5;



QY 4 GAGGCTGTGGTGAAGAGCTGGCATCCGSCCTGAGCGCAGCGGTGCGGTAGTTCG 63  
Db |||||  
QY 133 GAGGCTGTGGTGAAGAGCTGGCGATCCGSCCTGAGCGCAGCGGTGCGGTAGTTCG 192  
Db |||||  
QY 64 GCCAATGCGGCACCGCTGCTTCACACG-CGTTTTGCCGGGAGATGCGGCCGCTTCGTCC 122  
Db |||||  
QY 193 GCCAATGCGGCACCGCTGCTTCACACGCTGTTGTGGGAGATGCGGCCGCTTCGTCC 252  
Db |||||  
QY 123 TCTCAGTAAAGATGCGCGCTG-AGGCTGGGATTCAAATATCGCGTCAATAGA 181  
Db |||||  
QY 253 TCTCAGTCAAGACGCTGTCGTAGGACTGGATTTCAAATATCGCGTCAATAGA 312  
Db |||||  
QY 182 GAATGCTTTTCAATTCCTCCCAAGAAATCTGTCGTTTGTGGAACTGTGACAGA 241  
Db |||||  
QY 313 GAATGATTTTCAATTCCTCCCAAGAAATCTGTCGTTTGTGGAACTGTGACAGA 372  
Db |||||  
QY 242 AGCTTTGCTGAAGTACAAAGGGTGAACAAATGACTTTGAGTTGTTGAAGAACCAGCT 301  
Db |||||  
QY 373 AGCTTTGCTGAAGTACAAAGGGTGAACAAATGACTTTGAGTTGTTGAAGAACCAGCT 432  
Db |||||  
QY 302 GTTAGATCAGACATGAAGATGA-CCAGATCATCACT-GGCTGCTAGAAATCCGTTCT 359  
Db |||||  
QY 433 GTTAGATCAGACATGAAGATGA-CCAGATCATCACT-GGCTGCTAGAAATCCGTTCT 492  
Db |||||  
QY 360 TCTATCATGCTACTTGACAAAAGACTTTGAGCAACTTATCAGTATATATTAAGATGCG- 418  
Db |||||  
QY 493 TCTGTCATGCTACTTGACAAAAGACTTTGAGCAACTTATCAGTATATATTAAGATGCG 552  
Db |||||  
QY 419 TTGGTTGAATAGAGTCAACAGTAGTGAAGAGTATTTGGCTTTTCTGGTAATCTTGT 478  
Db |||||  
QY 553 TTGGTTGAATAGAGTCAACAGTAGTGAAGAGTATTTGGCTTTTCTGGTAATCTTGT 612  
Db |||||  
QY 479 ATCAGACAGACTGTTTCTCAGACGCTGCTCAGACATGATGCTTCCCATTTTGTGCC 538  
Db |||||  
QY 613 ATCAGACAGACTGTTTCTCAGACGCTGCTCAGACATGATGCTTCCCATTTTGTGCC 672  
Db |||||  
QY 539 TCCCCAGTATCATTAAGGAAGCGATGATGATGTTTCAGATTTCTGATCATGAAGATGA 598  
Db |||||  
QY 673 TCCCCAGTATCATTAAGGAAGCGATGATGATGTTTCAGATTTCTGATCATGAAGATGA 732  
Db |||||  
QY 599 TAATCTTCTGCAAAATTTTGACACATGTCACAGAGCTTGCAA 641  
Db |||||  
QY 733 TAATCTTCTGCAAAATTTTGACACATGTCACAGAGCTTGCAA 775  
Db |||||  
RESULT 10  
ABS64829  
ID ABS64829 standard; DNA; 2493 BP.  
AC ABS64829;  
XX  
DT 15-NOV-2002 (first entry)  
DE Human tumour suppressor gene #3.  
XX  
KW Human; cancer; gene; ds; tumour suppressor gene; breast cancer; lymphoma;  
KW adenocarcinoma; leukaemia; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW prostate; cervix; liver; ovary; adrenal gland; heart; brain; lung; colon;  
KW placenta; skeletal muscle; synovial membrane; tonsil; kidney; uterus;  
XX skin; cytostatic.  
OS Homo sapiens.  
XX  
PN WQ200264775-A1.  
XX  
PD 22-AUG-2002.  
XX  
PF 12-FEB-2002; 2002WO-AU000137.  
XX  
PR 12-FEB-2001; 2001AU-00003052.  
PR 12-FEB-2001; 2001AU-00003053.  
PR 12-FEB-2001; 2001AU-00003054.  
XX

(BION-) BIONOMICS LTD.  
PA Callen DF, Powell JA, Kremmidiotis G, Gardner AE, Whitmore SA;  
XX P-PSDB; ABG79517.  
DR WPI; 2002-657597/70.  
XX  
PT New tumor suppressor genes identified at 16q24.3, useful for identifying  
PT or obtaining full-length human genes involved in the tumorigenic process,  
PT or in diagnosing or treating cancer (e.g. breast or cervix cancer)  
PT through gene therapy.  
XX  
PS Claim 10; Page 76-79; 218pp; English.  
XX  
CC The invention relates to a tumour suppressor gene and the protein it  
CC encodes. The genes are useful for identifying and/or obtaining full-  
CC length human genes involved in the tumorigenic process. The genes the  
CC polypeptides and antibodies to the polypeptides are useful in diagnosing  
CC cancer, in establishing the prognosis of a patient diagnosed with cancer,  
CC or in treating cancer through gene therapy. In particular, the cancer is  
CC breast cancer. The drugs that restore tumour suppressor activity are  
CC useful for treating breast cancer or for manufacturing a medicament for  
CC the treatment of breast cancer. The cancer may also be adenocarcinoma,  
CC leukaemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, or  
CC cancer of the prostate, cervix, liver, ovary, adrenal gland, heart,  
CC brain, lung, placenta, skeletal muscle, synovial membrane, tonsil,  
CC kidney, colon, uterus or skin. This sequence represents a human tumour  
CC suppressor gene of the invention  
XX  
SQ Sequence 2493 BP; 554 A; 700 C; 761 G; 478 T; 0 U; 0 Other;  
XX  
Query Match 27.4%; Score 567; DB 6; Length 2493;  
Best Local Similarity 98.3%; Pred. No. 3.1e-157;  
Matches 573; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1196 GGGATTGCGAGAGGCAATTTTGGAAATCTCTGGAAAAAATTCGAGACCCAAAGTATCC 1255  
Db 711 GGGATTGCGAGAGGCAATTTTGGAAATCTCTGGAAAAAATTCGAGACCCAAAGTATCC 770  
QY 1256 TGCCATCATCAGCAGGCTGCTGGAATATATTTGGAAGCTTTTGGCAAGACTAAAT 1315  
Db 771 TGCCATCATCAGCAGGCTGCTGGAATATATTTGGAAGCTTTTGGCAAGACTAAAT 830  
QY 1316 TATTCCTCTTATTAATCTGTAATAATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATA 1375  
Db 831 TATTCCTCTTATTAATCTGTAATAATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATA 890  
QY 1376 CTTTAATAACCGAGATTCGGGAACAAAGGCAATTCGCGATGTTGCTCTCCATGACCAAT 1435  
Db 891 CTTTAATAACCGAGATTCGGGAACAAAGGCAATTCGCAATGTTGCTCTCCATGACCAAT 950  
QY 1436 TTACTCAGCTCCCAAGCTGTGTTCTACACCTTTGTTTATAGACACAGCAGCTTTTGAG 1495  
Db 951 TTACTCAGCTCCCAAGCTGTGTTCTACACCTTTGTTTATAGACACAGCAGCTTTTGAG 1010  
QY 1496 CGAAACCTGAAAGAGGTTTGCAGTATCTTCAGACTCTGAATTTTGGCGGATAGTAT 1555  
Db 1011 CGAAACCTGAAAGAGGTTTGCAGTATCTTCAGACTCTGAATTTTGGCGGATAGTAT 1070  
QY 1556 GAGCCAGCTAAATCCCTCGAAGATTTGCCCTCAGTGGTTAACTTTTGTCTGCAAT 1615  
Db 1071 GAGCCAGCTAAATCCCTCGAAGATTTGCCCTCAGTGGTTAACTTTTGTCTGCAAT 1130  
QY 1616 CACAAATAGTACAGCTCGTCTCTCTGCTACACCATCATTCAGAGGAAACAATCGCCAGAT 1675  
Db 1131 CACAAATAGTACAGCTCGTCTCTCTGCTACACCATCATTCAGAGGAAACAATCGCCAGAT 1190  
QY 1676 GCTGCCAGTCAATTAGGATACCCCTGGAGAGACTCAGTGCAGATCTGCGACAAACCCGCT 1735  
Db 1191 GCTGCCAGTCAATTAGGATACCCCTGGAGAGACTCAGTGCAGATCTGCGACAAACCCGCT 1250  
QY 1736 GGCACCTTCTTCCCTTTGATCCCTGTGTGCTGAGAGCTCA 1778  
Db |||||



Db	1251	GGACACCTTCGTCCCTTTTGATCCCTGTGTGCTGAAGAGGTGA	1293
RESULT 11			
ADA52976			
ID	ADA52976	standard; cDNA; 2493 BP.	
XX			
AC	ADA52976;		
XX			
DT	20-NOV-2003	(first entry)	
XX			
DE	Human coding sequence, SEQ ID 544.		
XX			
KW	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;		
KW	Gene Therapy; human; secretory protein; membrane proteins; cancer;		
KW	Inflammatory disease; osteoporosis; neurological disease; gene; ss.		
OS	Homo sapiens.		
XX			
PN	EP1293569-A2.		
XX			
PD	19-MAR-2003.		
XX			
PF	21-MAR-2002; 2002EP-00006586.		
XX			
PR	14-SEP-2001; 2001JP-00328381.		
PR	24-JAN-2002; 2002US-0350435P.		
XX			
PA	(HELI-) HELIX RES INST.		
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.		
XX			
PI	Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;		
PI	Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;		
PI	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;		
XX			
DR	WPI; 2003-395539/38.		
DR	P-PSDB; ADA54615.		
XX			
PT	New polynucleotides encoding full-length polypeptides, e.g. secretory		
PT	and/or membrane proteins, useful for developing medicines for diseases in		
PT	which the gene is involved, or as target molecules for gene therapy.		
XX			
PS	Claim 1; SEQ ID NO 544; 205pp; English.		
XX			
CC	The present invention relates to novel human secretory or membrane		
CC	proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-		
CC	ADA54071). The coding sequences are useful in the gene therapy of		
CC	diseases caused by abnormalities of the proteins, e.g. cancer,		
CC	inflammatory diseases, osteoporosis or neurological disease.		
XX			
SQ	Sequence 2493 BP; 554 A; 700 C; 761 G; 478 T; 0 U; 0 Other;		
	Query Match 27.4%; Score 567; DB 7; Length 2493;		
	Best Local Similarity 98.3%; Pred. No. 3.1e-157;		
	Matches 573; Conservative 0; Mismatches 10; Indels 0; Gaps 0;		
QY	1196	GGGATTCGACAGGCGATTTTGGAAACATCTCTGGAAAAAATGCAGGCCCAAGTAATCC	1255
Db	711	GGGATTCGACAGGCGATTTTGGAAACATCTTTGGAAAAAATGCAGGATCCAAGTAATCC	770
QY	1256	TGCCATCATCAGGCGCTGCGAAATATATTCGAGCCTTTTGGCAAGAGCTTAATTT	1315
Db	771	TGCCATCATCAGGCGCTGCGAAATATATTCGAGCCTTTTGGCAAGAGCTTAATTT	830
QY	1316	TATTCTCTTATTACTGTAAATCATGCTAGATCTTTTGGTTAACTGGCTGCACATATA	1375
Db	831	TATTCTCTTATTACTGTAAACCATGCTAGATCTTTTGGTTAACTGGCTGCACATATA	890
QY	1376	CCTTAATACACAGGATTCGGAAACAAAGGCATTCGCGATGTGCTTCCATGGACCAAT	1435
Db	891	CCTTAATACACAGGATTCGGAAACAAAGGCATTCGCAATGTGCTTCCATGGACCAAT	950
QY	1436	TTACTCAGCCTGCCAAGCTGTCTTACACCTTTGTTTTAGACACAGAGAGCTTTTGAG	1495
Db	951	TTACTCAGCCTGCCAAGCTGTCTTACACCTTTGTTTTAGACACAGAGCTTTTGAG	1010
QY	1496	CGGAACCTGAAAGAGGTTTGCAGTATCTTTCAGAGTCTGAAATTTTGAGCGGATATGAT	1555
Db	1011	CGGAACCTGAAAGAGGTTTGCAGTATCTTTCAGAGTCTGAAATTTTGAGCGGATATGAT	1070
QY	1556	GAGCCAGCTAAATCCCTGAAGATTGCTGCCTCAGTGGTTAACTTTTGGCTGCAAT	1615
Db	1071	GAGCCAGCTAAATCCCTGAAGATTGCTGCCTCAGTGGTTAACTTTTGGCTGCAAT	1130
QY	1616	CAAAATAAGTACAGCTCGTCTTCTGTACACCATCATTCAGAGGAAACAATGCCAGAT	1675
Db	1131	CAAAATAAGTACAGCTCGTCTTCTGTACACCATCATTCAGAGGAAACAATGCCAGAT	1190
QY	1676	GCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCGCT	1735
Db	1191	GCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCGCT	1250
QY	1736	GGACACCTTCTTCCCTTTTGATCCCTGTGTCTGAAGAGGTCA	1778
Db	1251	GGACACCTTCTTCCCTTTTGATCCCTGTGTCTGAAGAGGTCA	1293
RESULT 12			
AAS87643			
ID	AAS87643	standard; cDNA; 1260 BP.	
XX			
AC	AAS87643;		
XX			
DT	13-FEB-2002	(first entry)	
XX			
DE	DNA encoding novel human diagnostic protein #23447.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
PF	30-MAR-2001; 2001WO-US008631.		
XX			
PR	31-MAR-2000; 2000US-00540217.		
PR	23-AUG-2000; 2000US-00649167.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Drmanac RT, Liu C, Tang YT;		
XX			
DR	WPI; 2001-639362/73.		
DR	P-PSDB; ABG23456.		
XX			
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity.		
XX			
PS	Claim 1; SEQ ID NO 23447; 103pp; English.		
XX			
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)		
CC	sequences. (I) is useful as hybridisation probes, polymerase chain		
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,		
CC	and in recombinant production of (II). The polynucleotides are also used		
CC	in diagnostics as expressed sequence tags for identifying expressed		
CC	genes. (I) is useful in gene therapy techniques to restore normal		
CC	activity of (II) or to treat disease states involving (II). (II) is		
CC	useful for generating antibodies against it, detecting or quantitating a		
CC	polypeptide in tissue, as molecular weight markers and as a food		
CC	supplement. (II) and its binding partners are useful in medical imaging		
CC	of sites expressing (II). (I) and (II) are useful for treating disorders		

CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 1260 BP; 344 A; 227 C; 269 G; 420 T; 0 U; 0 Other;

Query Match 26.6%; Score 550.8; DB 5; Length 1260;  
Best Local Similarity 89.0%; Pred. No. 1.4e-152;  
Matches 696; Conservative 0; Mismatches 77; Indels 9; Gaps 9;

QY 4 GAGGCTGTGGCTGGAAGGAGCTGGGCATCGGCTCAGGCGAGCGGTGCGGTAGTTCG 63  
DB 133 GAGGCTGTGGCTGGAAGGAGCTGGGCATCGGCTCAGGCGAGCGGTGCGGTAGTTCG 192  
QY 64 GCCAATGGCGGCACCGCTGCTTCACAG -CGTTTCCGGGAGATCGGCGCGCTTCGTC 122  
DB 193 GCCAATGGCGGCACCGCTGCTTCACACGTTGTTCTCGGAGATCGGCGCGCTTCGTC 252  
QY 123 TCTGCAATTAAGAGCTGGCGGCTCG -AGGACTGGGATTCAAATATCGTGCATTAGA 181  
DB 253 TCTGCAATTAAGAGCTGGCGGCTCGGCTGAGGATTCGAAATATCGTGCATTAGA 312  
QY 182 GAATGACTTTTCAATTTCCCCCAAGAAAACTGTTGGTGGAACTGTGACAGA 241  
DB 313 GAATGATTTTCAATTTCCCCCAAGAAAACTGTTGGTGGAACTGTGACAGA 372  
QY 242 AGTCTTGTGAAGTACAAAAGGGTGAACAATGACTTTGAGTTGTTGAAGAACAGCT 301  
DB 373 AGTCTTGTGAAGTACAAAAGGGTGAACAATGACTTTGAGTTGTTGAAGAACAGCT 432  
QY 302 GTTAGATCCAGACATTAAGATGA -CCAGATCATCAACT -GGCTGCTAGAAATCCGTTCT 359  
DB 433 GTTAGATCCAGACATTAAGATGATCCAGATCATCAACTGGGCTGTAGAAATCCGTTCT 492  
QY 360 TCTATCATGTACTTGA -CAAAAGACTTTGAG -CAACTTATCATATTATTAAGATTGC 417  
DB 493 TCTGTATGTACTTGACCAAAAGACTTTGAGCAACTTATCATATTATTAAGATTGC 552  
QY 418 C-TTGGTTGAATGAAAGTC -AAACAGTAGTGAAGAGTATTTGGCTTTCTTCGTAATCT 475  
DB 553 CTTTGGTTGAATGAAAGTC -AAACAGTAGTGAAGAGTATTTGGCTTTCTTCGTAATCT 612  
QY 476 TGTATCAGCAGACAGCTTTTCTCAGACCGTGTCTCAGCATGATGCTT -CCATTTTG 534  
DB 613 TGTATCAGCAGACAGCTTTTCTCAGACCGTGTCTCAGCATGATGCTTCCCATTTTG 672  
QY 535 TGCTCTCCCGAGTGATCATTTAAGAGCGGATGTAGATGTTTCAGATTTCTGATCAAG 594  
DB 673 TGCTCTCCCGAGTGATCATTTAAGAGCGGATGTAGATGTTTCAGATTTCTGATCAAG 732  
QY 595 ATGATATCTCTCTGCAAAATTTGACACATGTCACAGAGCTTGCAAAATATGACAGAT 654  
DB 733 ATGATATCTCTCTGCAAAATTTGACACATGTCACAGAGCTTGCAAAATATGACAGAT 792  
QY 655 ATGTACCATCGACACCGTGGTTTCTCATGCCAATCTGTTGGGAAAAATTTCCATTTGTC 714  
DB 793 ATGTACCATCATGATGATATCTTTTCCCTTATTTGAATGTTTAATCTCAAGAAAAATGTA 852  
QY 715 GAAATATCAGAGAGAACACTGGAATGTTACGTTTCATTAATCTTACTAAGATTAGTATATT 774  
DB 853 ATCAATTTAGTAAATAATTAATAATGTTAATAGTATTAAGCTTGAGTCTTACATTTGCAAT 912  
QY 775 TT 776  
DB 913 TT 914

RESULT 13

AAS80955/C  
ID AAS80955 standard; cDNA; 655 BP.

XX AC AAS80955;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #16759.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PT 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DP P-PSDB; ABG16768.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.

PS Claim 1; SEQ ID NO 16759; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used  
XX in diagnostics as expressed sequence tags for identifying expressed  
XX genes. (I) is useful in gene therapy techniques to restore normal  
XX activity of (II) or to treat disease states involving (II). (II) is  
XX useful for generating antibodies against it, detecting or quantitating a  
XX polypeptide in tissue, as molecular weight markers and as a food  
XX supplement. (II) and its binding partners are useful in medical imaging  
XX of sites expressing (II). (I) and (II) are useful for treating disorders  
XX involving aberrant protein expression or biological activity. The  
XX polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
XX coding sequences of the invention. Note: The sequence data for this  
XX patent did not appear in the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 655 BP; 207 A; 134 C; 115 G; 199 T; 0 U; 0 Other;

Query Match 23.1%; Score 477.8; DB 5; Length 655;

Best Local Similarity 90.2%; Pred. No. 4.8e-131;

Matches 578; Conservative 0; Mismatches 57; Indels 6; Gaps 6;

QY 108 GCGGCGGCTTCGTCCTCTGCAAGTTAAGAGCTGGGCGGTCG-AGGACTGGATTTCAAA 166

DB 655 GCGGCGGCTTCGTCCTCTGCAAGCTGTCGCGTCGTAGGACTGGATTTCAAA 596

QY 167 TATCGGTGCAATTAGAGATGACTTTTCAATTTCCCCCAAGAAAACTGTCGTTGG 226

Db 595 TATGCGTCATTAGAGAAATGATTTTCAATCTCCCCAGAAAACTGTTGCGTTGG 536  
QY 227 TGAACCTGTGACAGAAAGTCTTGTGAAGTACAAAAAGGTGAAACAAATGACTTTGAGTT 286  
Db 535 TGGAACTGTGACAGAAAGTCTTGTGAAGTACAAAAAGGTGAAACAAATGACTTTGAGTT 476  
QY 287 GTTGAAGAACCCAGCTGTAGATCCGACATAAAGATGA-CCAGATCACTCACTGGCTGC 345  
Db 475 GTTGAAGAACCCAGCTGTAGATCCGACATAAAGATGACCCAGATCACTCACTGGCTGC 416  
QY 346 TAGAATCCGTTCTTCTATCATGTACTTGA-CAAAAGACTTTGAG-CAACTTATCAGTAT 403  
Db 415 TAGAATCCGTTCTTCTGTGTCATGTACTTGCACAAAGACTTTGAGCAACTTATCAGTAT 356  
QY 404 TATATTAAGATTCGCTTGTGTTGAATGAAGTGC-AAACAGTAGTGAAGAGTATTTGGCTT 462  
Db 355 TATATTGAGATTCGCTTGTGTTGAATGAAGTGC-AAACAGTAGTGAAGAGTATTTGGCTT 296  
QY 463 TTTCTTGGTAATCTGTATCAGACACACGCTTTTCTCAGACGGTGTCTCAGCATGATTG 522  
Db 295 TTCCTTGGTAATCTGTATCAGCATAGACTGTTTCTCAGACCGTGTCTCAGCATGATTG 236  
QY 523 CTT-CCCATTTTGTGCTCCCGAGTGATCAATTAAGGAAGCGGATGTAGATGTTTCAGAT 581  
Db 235 CTTCCCATTTTGTGCTCCCGAGTGATCAATTAAGGAAGCGGATGTAGATGTTTCAGAT 176  
QY 582 TCTGATGATGAAGATGATAATCTTCTGCAAAATTTTGACATGTCACAGAGCCTTGCAA 641  
Db 175 TCTGATGATGAAGATGATAATCTTCTGCAAAATTTTGACATGTCACAGAGCCTTGCAA 116  
QY 642 ATAATAGCAAGATATGTACCATCGACACCGTGTCTCAGTCCCAATCTGTTGGAAGAA 701  
Db 115 ATAATAGCAAGATATGTACCATCGATGATGTTTCTTCTTATTTTGAATGTTTAAATCT 56  
QY 702 TTTCCATTTGTTGCAAAATCAGAGAAACACTGGAATGTTA 742  
Db 55 CAAGAAATTTGTAATCAATTAGTAAATTTAATAATGTTA 15

## RESULT 14

AAS87641

ID AAS87641 standard; cDNA; 3169 BP.

XX AC AAS87641;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #23445.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-FSDB; ABG23454.

XX FT New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

XX Claim 1; SEQ ID NO 23445; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activities. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 3169 BP; 812 A; 841 C; 779 G; 737 T; 0 U; 0 Other;

QY Query Match 22.5%; Score 466.2; DB 5; Length 3169;

Db Best Local Similarity 64.6%; Pred. No. 3.5e-127;

QY Matches 933; Conservative 0; Mismatches 213; Indels 299; Gaps 4;

QY 69 ATGGCGGACCGCTCTTCACACGCGTTTGGCGGAGATGCGCGCTTCGCTCTGCA 128

Db 1 ATGGCGGACCGCTCTTCACACGCGTTTGGCGGAGATGCGCGCTTCGCTCTGCA 60

QY 129 GTTAAGAGCTGGGCGCTCGAGGACTGGGATTTCAAAATATGCGTGCATTAGAGATGAC 188

Db 61 GTCAAGATGCTGGGCGCTCGAGGACTGGGATTTCAAAATATGCGTGCATTAGAGATGAT 120

QY 189 TTTTTCATTTCCCGCCCAAGAAAACCTGTCGTTTGGTGAACCTGTGACAGAGCTTTG 248

Db 121 TTTTTCATTTCCCGCCCAAGAAAACCTGTCGTTTGGTGAACCTGTGACAGAGCTTTG 180

QY 249 CTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTTTGAAGAACCTGTGATAGAT 308

Db 181 CTGAAGTACAAAAGGGTGAACAAATGACTTTGAGTTTGAAGAACCTGTGATAGAT 240

QY 309 CCAGACATAAAGAGTGCATCAACTGGCTGCTAGAAATTCGTTCTTCTATCATG 368

Db 241 CCAGACATAAAGAGTGCATCAACTGGCTGCTAGAAATTCGTTCTTCTATCATG 300

QY 369 TACTTGACAAAGAGTGTGACCAACTTATCAGTATTAATTAAGATTCGTTGTTGAAT 428

Db 301 GTGGCGAGTGTGCTTTTCATCCCTTCGTTGCTTTCAGTTGCTCAAAACAGAGTGGC 360

QY 429 AGAAGTCAAAACAGTAGTGAAGAGTATTTGCGCTTTTCTGTTATCTTGTATCAGCAG 488

Db 361 CTAGCCGAGAGTGTAGCTG-----CTGGGCTGGGTCG 399

QY 489 ACTGTTTTCCTCAGACCGTGTCTCAGCATGATTTCCCATTTTGTGCTCCCGGAGTG 548

Db 400 CCCGTGGCTTGGCGCCACATACAGCCCGCTCTGCCACACACCCCAACGCCCGCAGC 459

QY 549 ATCATTAAGAGGCGATGATAGTTTTCAGATTTCTGATGATGAAGATGATAATCTTCT 608

Db 460 CAGCTTCTGGCTCACTCATCTGCTTTGCA----- 488

QY 609 GCAAAATTTGACACATGTCACAGAGCCTTGCAATAATAGCAAGATATGTACCATGACA 668

Db 489 -----GGACA 493

QY 669 CCCTGGTTTCTCATGCCAATACCTGGTGGAAAAATTTCCATTTGTTGAAAAATCAGAGAGA 728

Db 494 CCGTGGTTCTCATGCCATACTGGTGGAAAATTTCCATTTGTCGAAAATCAGAGAGA 553  
QY 729 ACATGGAAATGTACGTTTCATCACTTCTAAGAGATAGTGTATATTTTCCAACTTGAGG 788  
Db 554 ACATGGAAATGTACGTTTCATCACTTCTAAGAGATAGTGTATATTTTCCAACTTGAGG 613  
QY 789 CATGAAATTTCTGAGCTTATTTTGAAGAACTACTCAAGTTGGATGTGAATCCCGG 848  
Db 614 CATGAAATTTCTGAGCTTATTTTGAAGAACTACTCAAGTTGGATGTGAATCCCGG 673  
QY 849 CAGGGTATTGAAGATCTCAAGAAACAGCAACTCAAACTTGTGGGACAGATTCACG 908  
Db 674 CAGGGTATTGAAGATCTCAAGAAACAGCAACTCAAACTTGTGGGACAGATTCACG 733  
QY 909 GAAGGATTTTAAATATGATGAAGATGAAGAACTGAACATGAACAAAGGCTGTCT 968  
Db 734 GAAGGATTTGTAAATATGG----- 752  
QY 969 GAACGCTCGACAGATGTGTATCTGTAGCCGAGCGCTGGACATCTGTATCTTTG 1028  
Db 753 ----- 752  
QY 1029 GTTTTGTCTCATGAAGATGTCTGTATGTAGATGTAAGTTGATACGCGCAAAACA 1088  
Db 753 --TTACAGCTTTAATGAAGTGGAGATGAAGTTTATCAATCAAAAGGTTGGAACA 810  
QY 1089 AAGGATCTATATCGGACCTGATAAACAATCTTTGACAAACTCTGTGTGCCACCATGCC 1148  
Db 811 G----- 811  
QY 1149 TCCTGCCATGTACAGTTTTTTCATGTTTACCTCTGTAGTTTCAAAATTTGGGATTCGCAGAG 1208  
Db 812 -----CTAGTGTCTCATCTTTTGGATTCGCAGAG 841  
QY 1209 GCATTTTTTGAACATCTCTGGAAGAAATTCGAGGACCCCAAGTAATCTGTCATCATCAGG 1268  
Db 842 GCATTTTTTGAACCTCTTTGGAAGAAATTCGAGGACCCCAAGTAATCTGTCATCATCAGG 901  
QY 1269 CAGGCTGTGGAAATTAATATGGAAGCTTTTGGGAGAGCTAAATTTATCTCTTATT 1328  
Db 902 CAGGCTGTGGAAATTAATATGGAAGCTTTTGGGAGAGCTAAATTTATCTCTTATT 961  
QY 1329 ACTGTAAATCATGCTCTAGATCTTTTGGTGTAACTGTGTCACATATACCTTAATAACCCAG 1388  
Db 962 ACTGTAAATCATGCTCTAGATCTTTTGGTGTAACTGTGTCACATATACCTTAATAACCCAG 1021  
QY 1389 GATTCGGGAACAAAGGCATCTCGATGTTGTCTCTCATGAGCCATTTTACTCAGCCTGC 1448  
Db 1022 GATTCGGGAACAAAGGCATCTCGATGTTGTCTCTCATGAGCCATTTTACTCAGCCTGC 1081  
QY 1449 CAAGCTGTGTTTACACCTTTGTTTGTAGACACAGCAGCTTTTGGCGGAAACCTTGAA 1508  
Db 1082 CAAGCTGTGTTTACACCTTTGTTTGTAGACACAGCAGCTTTTGGCGGAAACCTTGAA 1141  
QY 1509 GAAGG 1513  
Db 1142 GAAGG 1146

RESULT 15

ID ABA83038  
XX ABA83038 standard; DNA; 1461 BP.  
AC ABA83038;  
XX  
DT 05-FEB-2002 (first entry)  
DE Human transcription factor TRFX-65 coding sequence.  
XX Human; transcription factor; TRFX; cell proliferative disease;  
KW autoimmune disease; inflammation; neurological disease;  
KW developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;

neuroprotective; antiinflammatory; gene therapy; ds.  
XX Homo sapiens.  
OS WO200172777-A2.  
XX 04-OCT-2001.  
XX 13-MAR-2001; 2001WO-US008117.  
XX 13-MAR-2000; 2000US-0189886P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Hillman JL, Baughn MR, Yue H, Lal P, Lu DM, Patterson C;  
PI Azimzai Y, Bandman O, Tang YT, Mathur P, Shan P, Au-Young J;  
PI Reddy R;  
XX WPI; 2001-570896/64.  
DR P-PSDB; ABB50214.  
XX Novel transcription factor polypeptides, used to treat diseases  
PT associated with altered activity and expression of TRFX, and to screen  
PT for agents capable of modulating its activity.  
XX Claim 11; Page 299; 327pp; English.  
CC The present sequence is the coding sequence for a human transcription  
CC factor. The transcription factor and its coding sequence are useful in  
CC the diagnosis, treatment and prevention of diseases associated with  
CC altered expression of the transcription factor e.g. cell proliferative,  
CC autoimmune/inflammatory, neurological and developmental disorders. A  
CC number of specific disorders/diseases are given in the specification,  
CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,  
CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic  
CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,  
CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,  
CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative  
CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's  
CC disease, stroke, and viral, bacterial, fungal and protozoal infections  
XX  
SQ Sequence 1461 BP; 413 A; 335 C; 298 G; 415 T; 0 U; 0 Other;  
Query Match 20.3%; Score 420.8; DB 5; Length 1461;  
Best Local Similarity 97.3%; Pred. No. 7.1e-114;  
Matches 428; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 1188 TTCAAAATGGGATTCGAGAGGCAATTTTGGAAACATCTCTGGAAGAAATTTGCAAGCCCA 1247  
Db 530 TTATATATGGGATTCGAGAGGCAATTTTGGAAACATCTCTGGAAGAAATTTGCAAGCCCA 589  
QY 1248 AGTAATCCTGCCATCATCAGGAGCGCTGCTGGAATATATATTTGGAAGCTTTTGGCAAGA 1307  
Db 590 AGTAATCCTGCCATCATCAGGAGCGCTGCTGGAATATATATTTGGAAGCTTTTGGCAAGA 649  
QY 1308 GCTAAATTTATTTCTTATTTACTGTAATCATGCTAGATCTTTTGGTAACTGGCTG 1367  
Db 650 GCTAAATTTATTTCTTATTTACTGTAATCATGCTAGATCTTTTGGTAACTGGCTG 709  
QY 1368 CACATATACCTTAATTAACAGGATTCGGGAACAAAGGCATTCGCGATGTGCTCTCCAT 1427  
Db 710 CACATATACCTTAATTAACAGGATTCGGGAACAAAGGCATTCGCGATGTGCTCTCCAT 769  
QY 1428 GGACCAATTTTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAGCAG 1487  
Db 770 GGACCAATTTTACTCAGCTGCCAAGCTGTGTTCTACACCTTTGTTTGTAGACACAGCAG 829  
QY 1488 CTTTTCGAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAAATTTGAGCGG 1547  
Db 830 CTTTTCGAGCGGAAACCTGAAAGAGGTTTGCAGTATCTTCAGAGTCTGAAATTTGAGCGG 889  
QY 1548 ATAGTATGAGCGGAGCAATTAATCCCTGAAGATTTGCTGCTGCCCTCAGTGTAACTTTT 1607

Db 890 ATAGTGATGAGCCAGCTAAATCCCTGAAGATTGGCTGCGCTCAGTGGTTAACTTTTT 949

Qy 1608 GCTGCAATCACAATAAGTA 1627

|||||

Db 950 GCTGCAATCACAAGATGAA 969

Search completed: July 18, 2004, 02:37:31  
Job time : 827 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 02:18:21 ; Search time 153 Seconds  
(without alignments)  
7500.905 Million cell updates/sec

Title: US-09-932-678-1  
Perfect score: 2068  
Sequence: 1 acacaggctggctggaag.....tgacattggattccccat 2068

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	208.2	10.1	513	4	US-09-621-976-3086
2	48.2	2.3	8797	2	US-08-723-306-6
3	48.2	2.3	8797	5	PCT-US96-10041-6
4	48.2	2.3	11093	2	US-08-723-306-5
5	48.2	2.3	11093	5	PCT-US96-10041-5
C 6	43.4	2.1	399	4	US-09-621-976-8976
C 7	40.4	2.0	7218	1	US-08-232-463-14
8	39.2	1.9	3108	4	US-09-206-942-72
9	39.2	1.9	4937	1	US-08-038-682-3
10	39.2	1.9	4937	2	US-08-302-832-3
11	39.2	1.9	4937	2	US-08-550-198-3
12	39.2	1.9	4937	2	US-08-469-880-3
13	39.2	1.9	4937	2	US-08-728-470-3
14	39.2	1.9	4937	2	US-08-617-697-3
15	39.2	1.9	4937	3	US-08-719-641-3
16	39.2	1.9	4937	4	US-09-206-942-70
C 17	38	1.8	832	4	US-09-621-976-2813
C 18	37.8	1.8	2418	4	US-09-601-198-61
19	37.8	1.8	2910	4	US-09-206-942-31
20	37.8	1.8	2928	4	US-09-206-942-29
21	37.2	1.8	7518	4	US-09-620-312D-1051
22	36.6	1.8	1497	4	US-09-220-132-94
23	36.2	1.8	399	4	US-09-621-976-8976
24	36.2	1.8	586	4	US-09-669-751-55
25	35.8	1.7	832	4	US-09-621-976-2813
26	35.6	1.7	1398	4	US-09-134-001C-1019
C 27	35.6	1.7	2148	4	US-09-219-983A-19

C 28	35.6	1.7	2492	4	US-09-219-983A-18	Sequence 18, Appl
29	35.4	1.7	1461	4	US-09-369-247-49	Sequence 49, Appl
30	35.2	1.7	5538	2	US-08-231-193A-55	Sequence 55, Appl
31	35.2	1.7	5538	2	US-08-486-273A-55	Sequence 55, Appl
32	35.2	1.7	5538	3	US-08-940-086A-55	Sequence 55, Appl
33	35.2	1.7	5538	4	US-08-940-035A-55	Sequence 55, Appl
34	35.2	1.7	5538	4	US-08-935-105A-55	Sequence 55, Appl
35	35.2	1.7	5538	4	US-09-648-797-55	Sequence 55, Appl
36	35.2	1.7	5538	4	US-09-386-123-55	Sequence 55, Appl
37	35.2	1.7	5983	4	US-08-264-578-1	Sequence 1, Appl
38	35.2	1.7	10878	4	US-09-911-842A-1	Sequence 1, Appl
39	35	1.7	3757	4	US-09-620-312D-106	Sequence 106, App
C 40	35	1.7	1230025	4	US-09-313-294A-5811	Sequence 1, Appl
41	34.8	1.7	304	4	US-09-134-001C-2374	Sequence 2374, Ap
42	34.8	1.7	771	4	US-09-134-001C-2374	Sequence 714, App
43	34.8	1.7	978	4	US-09-433-699-3	Sequence 3, Appl
44	34.6	1.7	2518	3	US-09-206-942-48	Sequence 48, Appl
45	34.6	1.7	3222	4		

ALIGNMENTS

RESULT 1  
US-09-621-976-3086  
; Sequence 3086, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 3086  
; LENGTH: 513  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 25..231  
US-09-621-976-3086

Query Match	10.1%;	Score 208.2;	DB 4;	Length 513;
Best Local Similarity	98.6%;	Pred. No. 5.9e-54;		
Matches 210;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;
QY	1650	ATCATTGAGAGGAACAATCGCCAGATGCTGCCAGTCTATTAGGAGTACCGCTGGAGGAGAC	1709	
Db	1	ATCATTGAGAGGAACAATCGCCAGATGCTGCCAGTCTATTAGGAGTACCACTGGAGGAGAC	60	
QY	1710	TCAGTCAGATCTGCACAAACCCGCTGGACACCTTTTCCCTTTGATCCCTGTGTGCTG	1769	
Db	61	TCAGTCAGACCTGTCACAAACCCCACTGGACACCTTCTTCCCTTTGATCCCTGTGTGCTG	120	
QY	1770	AAGAGGTCGAAGAAATTCATTGATCCCTATTATCAGTATGGGAAGACATGAGTGTCTGAA	1829	
Db	121	AAGAGGTCGAAGAAATTCATTGATCCCTATTATCAGTATGGGAAGACATGAGTGTCTGAA	180	
QY	1830	GAGCTACAGGATTCAGAAACCCATGAAAG	1862	
Db	181	GAGCTACAGGATTCAGAAACCCATGAAAG	213	

RESULT 2  
US-08-723-306-6  
; Sequence 6, Application US/08723306  
; Patent No. 5856178  
; GENERAL INFORMATION:  
; APPLICANT: White PhD, Kenneth

APPLICANT: Morrey PhD, John  
APPLICANT: Reed, William  
TITLE OF INVENTION: Cassette for Expression of Lytic  
Peptides in Mammalian Transgenic Organisms  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Trask Britt and Rossa  
STREET: P.O. Box 2550  
CITY: Salt Lake City  
STATE: Utah  
COUNTRY: USA  
ZIP: 84110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/723,306  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Sweigert PhD, Susan E  
REGISTRATION NUMBER: 36,289  
REFERENCE/DOCKET NUMBER: 2549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 8015321922  
TELEFAX: 8015319168  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8797 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: not relevant  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Construct comprising Bos  
taurus beta casein 5' regulatory region plus genes encoding  
amphipathic peptide and green fluorescent protein"  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
PCT-US96-10041-6

Query Match 2.3%; Score 48.2; DB 2; Length 8797;  
Best Local Similarity 52.2%; Pred. No. 0.00083;  
Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 697 AAAAAATTCATTTGTTGAAAAATCAGAGAACACTGGAATGTTACGTTCACTTAC 756  
DB 2278 AAAGAAGAGTATATTTATTTAAATTTGCTCAGAACATCCAAATTTCAAGTTTATCATTTAT 2337  
QY 757 TAAGGATTAGTGTATATTTTCCAACTTGAGGCATGAATTTCTGGAGCTTATTTGAAA 816  
DB 2338 CTTACAATATTTCAAAAATTTAAATATAGATACATGAATACAGAAATTAATTAAGAGA 2397  
QY 817 AACTACTCAAGTTGGATGTGAATGCATCCGGCAGGATTTGAAGATGCTGAAGAACAG 876  
DB 2398 AGATATTTTATTTTGAATAAAATTTCTAGGTTGACAGGAGTACCAGGAACAAAA 2457  
QY 877 CAACCTCAACTGTGGTGGGACAGA 901  
DB 2458 ACAATGAAAAATGTGATCTGACAGA 2482

RESULT 3  
PCT-US96-10041-6  
Sequence 6, Application PC/TUS9610041  
GENERAL INFORMATION:  
APPLICANT: White PhD, Kenneth  
APPLICANT: Morrey PhD, John  
APPLICANT: Reed, William  
TITLE OF INVENTION: Cassette for Expression of Lytic  
Peptides in Mammalian Transgenic Organisms  
NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Trask Britt and Rossa  
STREET: P.O. Box 2550  
CITY: Salt Lake City  
STATE: Utah  
COUNTRY: USA  
ZIP: 84110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10041  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sweigert PhD, Susan E  
REGISTRATION NUMBER: 36,289  
REFERENCE/DOCKET NUMBER: 2549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 8015321922  
TELEFAX: 8015319168  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8797 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: not relevant  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Construct comprising Bos  
taurus beta casein 5' regulatory region plus genes encoding  
amphipathic peptide and green fluorescent protein"  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
PCT-US96-10041-6

Query Match 2.3%; Score 48.2; DB 5; Length 8797;  
Best Local Similarity 52.2%; Pred. No. 0.00083;  
Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 697 AAAAAATTCATTTGTTGAAAAATCAGAGAACACTGGAATGTTACGTTCACTTAC 756  
DB 2278 AAAGAAGAGTATATTTATTTAAATTTGCTCAGAACATCCAAATTTCAAGTTTATCATTTAT 2337  
QY 757 TAAGGATTAGTGTATATTTTCCAACTTGAGGCATGAATTTCTGGAGCTTATTTGAAA 816  
DB 2338 CTTACAATATTTCAAAAATTTAAATATAGATACATGAATACAGAAATTAATTAAGAGA 2397  
QY 817 AACTACTCAAGTTGGATGTGAATGCATCCGGCAGGATTTGAAGATGCTGAAGAACAG 876  
DB 2398 AGATATTTTATTTTGAATAAAATTTCTAGGTTGACAGGAGTACCAGGAACAAAA 2457  
QY 877 CAACCTCAACTGTGGTGGGACAGA 901  
DB 2458 ACAATGAAAAATGTGATCTGACAGA 2482

RESULT 4  
US-08-723-306-5  
Sequence 5, Application US/08723306  
Patent No. 5856178  
GENERAL INFORMATION:  
APPLICANT: White PhD, Kenneth  
APPLICANT: Morrey PhD, John  
APPLICANT: Reed, William  
TITLE OF INVENTION: Cassette for Expression of Lytic  
Peptides in Mammalian Transgenic Organisms  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Trask Britt and Rossa  
STREET: P.O. Box 2550  
CITY: Salt Lake City



```

; STATE: Utah
; COUNTRY: USA
; ZIP: 84110
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723,306
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sweigert Phd, Susan E
; REGISTRATION NUMBER: 36,289
; REFERENCE/DOCKET NUMBER: 2549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 8015321922
; TELEFAX: 8015319168
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11093 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Construct comprising
; portions of Bos taurus beta casein gene and genes encoding
; amphipathic peptide and green fluorescent protein"
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1801..1834
; OTHER INFORMATION: /product= "beta casein exon 1"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3780..3832
; OTHER INFORMATION: /product= "beta casein exon 2"
; FEATURE:
; NAME/KEY: TATA signal
; LOCATION: 1766..1773
; FEATURE:
; NAME/KEY: exon
; LOCATION: 4567..4590
; OTHER INFORMATION: /product= "portion of beta casein
; OTHER INFORMATION: exon 3"
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 127..1800
; OTHER INFORMATION: /function= "5' flanking regulatory
; OTHER INFORMATION: region of bovine beta casein gene"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4587..5310
; OTHER INFORMATION: /product= "Green fluorescent
; OTHER INFORMATION: protein"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 5320..5449
; OTHER INFORMATION: /product= "Shiva-1 coding sequence"
; FEATURE:
; NAME/KEY: polyA signal
; LOCATION: 7630..7635
; OTHER INFORMATION: /standard_name= "Bovine beta casein
; OTHER INFORMATION: 3' region, in exon 9"
; US-08-723-306-5
; Query Match 2.3%; Score 48.2; DB 2; Length 11093;
; Best Local Similarity 52.2%; Pred. No. 0.00096;
; Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
; 697 AAAAAATTCCTATTTGTCGAAAAATCAGAGAACACTGGGAATGTTACGTTCAACTTAC 756

```

```

; STATE: Utah
; COUNTRY: USA
; ZIP: 84110
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723,306
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sweigert Phd, Susan E
; REGISTRATION NUMBER: 36,289
; REFERENCE/DOCKET NUMBER: 2549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 8015321922
; TELEFAX: 8015319168
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11093 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Construct comprising
; portions of Bos taurus beta casein gene and genes encoding
; amphipathic peptide and green fluorescent protein"
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1801..1834
; OTHER INFORMATION: /product= "beta casein exon 1"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3780..3832
; OTHER INFORMATION: /product= "beta casein exon 2"
; FEATURE:
; NAME/KEY: TATA signal
; LOCATION: 1766..1773
; FEATURE:
; NAME/KEY: exon
; LOCATION: 4567..4590
; OTHER INFORMATION: /product= "portion of beta casein
; OTHER INFORMATION: exon 3"
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 127..1800
; OTHER INFORMATION: /function= "5' flanking regulatory
; OTHER INFORMATION: region of bovine beta casein gene"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4587..5310
; OTHER INFORMATION: /product= "Green fluorescent
; OTHER INFORMATION: protein"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 5320..5449
; OTHER INFORMATION: /product= "Shiva-1 coding sequence"
; FEATURE:
; NAME/KEY: polyA signal
; LOCATION: 7630..7635
; OTHER INFORMATION: /standard_name= "Bovine beta casein
; OTHER INFORMATION: 3' region, in exon 9"
; US-08-723-306-5
; Query Match 2.3%; Score 48.2; DB 2; Length 11093;
; Best Local Similarity 52.2%; Pred. No. 0.00096;
; Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
; 697 AAAAAATTCCTATTTGTCGAAAAATCAGAGAACACTGGGAATGTTACGTTCAACTTAC 756

```

## RESULT 5

PCT-US96-10041-5

; Sequence 5, Application PC/TUS9610041

; GENERAL INFORMATION:

; APPLICANT: White Phd, Kenneth

; APPLICANT: Morrey Phd, John

; APPLICANT: Reed, William

; TITLE OF INVENTION: Cassette for Expression of Lytic

; TITLE OF INVENTION: Peptides in Mammalian Transgenic Organisms

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Trask Britt and Rossa

; STREET: P.O. Box 2550

; CITY: Salt Lake City

; STATE: Utah

; COUNTRY: USA

; ZIP: 84110

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US96/10041

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Sweigert Phd, Susan E

; REGISTRATION NUMBER: 36,289

; REFERENCE/DOCKET NUMBER: 2549

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 8015321922

; TELEFAX: 8015319168

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 11093 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: not relevant

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "Construct comprising

; DESCRIPTION: portions of Bos taurus beta casein gene and genes encoding

; DESCRIPTION: amphipathic peptide and green fluorescent protein"

; HYPOTHETICAL: YES

; ANTI-SENSE: NO

; FEATURE:

; NAME/KEY: exon

; LOCATION: 1801..1834

; OTHER INFORMATION: /product= "beta casein exon 1"

; FEATURE:

; NAME/KEY: exon

; LOCATION: 3780..3832

; OTHER INFORMATION: /product= "beta casein exon 2"

; FEATURE:

; NAME/KEY: TATA signal

; LOCATION: 1766..1773

; FEATURE:

NAME/KEY: exon  
 LOCATION: 4567..4590  
 OTHER INFORMATION: /product= "portion of beta casein"  
 OTHER INFORMATION: exon 3"  
 FEATURE:  
 NAME/KEY: 5'UTR  
 LOCATION: 127..1800  
 OTHER INFORMATION: /function= "5' flanking regulatory  
 OTHER INFORMATION: region of bovine beta casein gene"  
 FEATURE:  
 NAME/KEY: mat peptide  
 LOCATION: 4587..5310  
 OTHER INFORMATION: /product= "Green fluorescent  
 OTHER INFORMATION: protein"  
 FEATURE:  
 NAME/KEY: mat peptide  
 LOCATION: 5320..5449  
 OTHER INFORMATION: /product= "Shiva-1 coding sequence"  
 FEATURE:  
 NAME/KEY: polyA signal  
 LOCATION: 7630..7635  
 OTHER INFORMATION: /standard name= "Bovine beta casein"  
 OTHER INFORMATION: 3' region, in exon 9"  
 PCT-US96-10041-5

Query Match 2.3%; Score 48.2; DB 5; Length 11093;  
 Best Local Similarity 52.2%; Pred. No. 0.00096;  
 Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 697 AAAAATTCATTTGTCGAAATCAGAGAGAACTGGAATGTACGTTCACTAC 756  
 DB 2278 AAAGAAGAGTATATTATTAAATTCCTCAGAACTCCAAATTTCAAGTTTATCATTTAT 2337  
 QY 757 TAAGGATAGTGTATATTTTCCAACTTGAGGCATGAAATCTGGAGCTTATTATTGAA 816  
 DB 2338 CTTACAATTTCAAAATATTAAATAGATACATGAATACAGAAGTAAATTAAGAGA 2397  
 QY 817 AACTACTCAAGTTGGATGTGAATCATCCCGCAGGATTTGAAGATGTGAAGAAACAG 876  
 DB 2398 AAGTATTTTATTGTAATAAAATTTCTAGTTGACAGGAGTACAGGAACAAAA 2457  
 QY 877 CAACTCAAACTTGCTGGTGACAGA 901  
 DB 2458 ACAATGAAAATGTGATCTGACAGA 2482

RESULT 6  
 US-09-621-976-8976/c  
 ; Sequence 8976, Application US/09621976  
 ; Patent No. 6639063  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, J.B.  
 ; APPLICANT: Jobert, S.  
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 ; FILE REFERENCE: GENSET.054PR2  
 ; CURRENT APPLICATION NUMBER: US/09/621.976  
 ; CURRENT FILING DATE: 2000-07-21  
 ; NUMBER OF SEQ ID NOS: 19335  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 8976  
 ; LENGTH: 399  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-621-976-8976

Query Match 2.1%; Score 43.4; DB 4; Length 399;  
 Best Local Similarity 11.8%; Pred. No. 0.00035;  
 Matches 35; Conservative 138; Mismatches 124; Indels 0; Gaps 0;

QY 895 GGACAGATCCACGAGAGGATTTTAAATGATGAAGATGAAGAACTGAACATGAAA 954  
 DB 330 SRGSRRYKTSAMWGRAKKKKTCKMYWKKGGKGSITYMAMRSRSGTGRWSYRRA 271

QY 955 CAAAGGCTGCTCTGAACCGCTCGACGATGTCATCTCTAGCGAGCGCTGACGA 1014  
 DB 270 MWRGSKSGGSGYRMAGYRSSRWSYSAMWKKKMTCKWGRSSWGRSTGYIYAWYKK 211  
 QY 1015 TCCTGATGCTCTTGGTTTGTCTCATACATGAAGATGTCTGCTATGTAGTGAAGTTG 1074  
 DB 210 SWCTSRKMYKKRRKKRRKCTSTKRTCYRGSTYKCAKYTKRKKWTRWYIYKSY 151  
 QY 1075 ATAAACGGCAAAACAAAGGATCTATATCGGACCTGATAACATCTTTGACAACTCTGT 1134  
 DB 150 MSKKTRWMTAYYTKRWKMTKTKWTCTCCKCTTYWAGTMMYRYRYIYAKRW 91  
 QY 1135 TGCCACCCATGCTCTCCATGACAGTTTTCATGTTTACCTCTCTAGTTTCA 1191  
 DB 90 SKRCTWSTTCYCKMYMAKCKWSYMSMSMKWKKSMWKKTYIYIYMMKSKMTYW 34

RESULT 7  
 US-08-232-463-14/c  
 ; Sequence 14, Application US/08232463  
 ; Patent No. 5670367  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DORNER, F.  
 ; APPLICANT: SCHEIFLINGER, F.  
 ; APPLICANT: FALKNER, F. G.  
 ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
 ; NUMBER OF SEQUENCES: 52  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 1800 Diagonal Road, Suite 500  
 ; CITY: Alexandria  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22313-0299  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/232.463  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/935.313  
 ; FILING DATE:  
 ; APPLICATION NUMBER: EP 91 114 300.6  
 ; FILING DATE: 26-AUG-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703)836-9300  
 ; TELEFAX: (703)683-4109  
 ; TELEX: 899149  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 7218 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; CLONE: pTZgpt-F15  
 US-08-232-463-14

Query Match 2.0%; Score 40.4; DB 1; Length 7218;  
 Best Local Similarity 6.8%; Pred. No. 0.19;  
 Matches 29; Conservative 207; Mismatches 188; Indels 0; Gaps 0;  
 QY 584 TGAATGATGAAGATGATAATCTTCTGCAATTTTGACACATGTACAGAGCCTTGCAAT 643

```
Db 1458 TAAAGATAGAGAAATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1399
Qy 644 AATAGCAAGATATACATCGACACCGTGCTTCTCATGCAATACTGTTGGAATAATT 703
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1398 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1339
Qy 704 TCCATTGTTGCAAAATCAGAGAAACACTGGAATGTTACATCTTCACTTACTAAGAT 763
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1338 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1279
Qy 764 TAGGTATATTTCCAACTTGGAGCATGAATTCGAGCTTATTATGAAAACACTACT 823
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1278 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1219
Qy 824 CAAGTTGATGTGAATCCCGCAGGCTGATGAGATGCTGAAGAAACAGCAACTCA 883
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1218 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1159
Qy 884 AACTTGTGGGACAGATTCACGGAAGGATGTTTAAATGATGAAGATGAAGAAAC 943
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1158 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1099
Qy 944 TGAACATGAACAAAGGCTGCTCTGACGGCTCGACAGATGGTGATCCTGTAGCGA 1003
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1098 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1039
Qy 1004 GCGC 1007
    |||
Db 1038 GCTC 1035

RESULT 8
US-09-206-942-72
; Sequence 72, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72
; LENGTH: 3108
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-09-206-942-72

Query Match 1.9%; Score 39.2; DB 4; Length 3108;
Best Local Similarity 53.2%; Pred. No. 0.26;
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 803 GCTTATTATTCGAAATCTACTCAAGTTGGATGTAATGTCATCCGCGCAGGCTATTGAAGA 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2613 GGTATTACGCAAAAGATGCTAACTAAATGGTGATGCATCAGTGATAGTACAGAAGT 2672
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 863 TGCTGAAGAAACAGCAACTCAAACCTTGTTGGGACAGATTCACGGAAGGATTGTTTAA 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2673 GAATGCACTCAACGCAAGCGCTCTGGTAGTGTGACTGGCGCAACTCAAGCAGTGTGAA 2732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 923 TATGATGAAGATGAAGAACTGAACATGAACAAA 958
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2733 TATCACTGGGATTATAACACAGATTAATGGGTTAA 2768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-08-038-682-3
; Sequence 3, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-038-682-3

Query Match 1.9%; Score 39.2; DB 1; Length 4937;
Best Local Similarity 53.2%; Pred. No. 0.34;
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 803 GCTTATTATTCGAAATCTACTCAAGTTGGATGTAATGTCATCCGCGCAGGCTATTGAAGA 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4287 GGTATTACGCAAAAGATGCTAACTAAATGGTGATGCATCAGTGATAGTACAGAAGT 4346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 863 TGCTGAAGAAACAGCAACTCAAACCTTGTTGGGACAGATTCACGGAAGGATTGTTTAA 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4347 GAATGCACTCAACGCAAGCGCTCTGGTAGTGTGACTGGCGCAACTCAAGCAGTGTGAA 4406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-08-302-832-3
; Sequence 3, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
```

COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/302,832  
FILING DATE: 16-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US pct/us93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstreser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-404  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4937 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-302-832-3

Query Match  
Best Local Similarity 53.2%; Score 39.2; DB 1; Length 4937;  
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 803 GCTTATTATTGAAAACTACTCAAGTTGGATGTAATGTCATCCCGCAGGGTATTGAAGA 862  
Db 4287 GGTATTACGCAAAAGATGCTAAGCTAAATGGTATGATCAGGTGATAGTACAGAACT 4346  
QY 863 TCGTGAAGAAACAGCAACTCAAACCTTGTTGGTGGGACAGATTCACGGAAGGATTGTTAA 922  
Db 4347 GAATGCAGTCAACGCAAGCGGCTCTGTTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406  
QY 923 TATGGATGAAGTGAAGAACTGAACATCAACAA 958  
Db 4407 TATCACTGGGGATTAAACACAGTAATGGGTAAA 4442

RESULT 11  
US-08-530-198-3  
Sequence 3, Application US/08530198  
Patent No. 5869065  
GENERAL INFORMATION:  
APPLICANT: BARENKAMP, STEPHEN J  
APPLICANT: ST. GEME III, JOSEPH W  
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESS: Shoemaker and Mattare, Ltd  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/530,198  
FILING DATE: 13-DEC-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: BERKSTRESSER, JERRY W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: JWB-1186  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4937 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-530-198-3

Query Match  
Best Local Similarity 53.2%; Score 39.2; DB 2; Length 4937;  
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 803 GCTTATTATTGAAAACTACTCAAGTTGGATGTAATGTCATCCCGCAGGGTATTGAAGA 862  
Db 4287 GGTATTACGCAAAAGATGCTAAGCTAAATGGTATGATCAGGTGATAGTACAGAACT 4346  
QY 863 TCGTGAAGAAACAGCAACTCAAACCTTGTTGGTGGGACAGATTCACGGAAGGATTGTTAA 922  
Db 4347 GAATGCAGTCAACGCAAGCGGCTCTGTTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406  
QY 923 TATGGATGAAGTGAAGAACTGAACATCAACAA 958  
Db 4407 TATCACTGGGGATTAAACACAGTAATGGGTAAA 4442

RESULT 12  
US-08-469-880-3  
Sequence 3, Application US/08469880  
Patent No. 5876733  
GENERAL INFORMATION:  
APPLICANT: BARENKAMP, STEPHEN J.  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESS: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,880  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:

```
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-880-3

Query Match
Best Local Similarity 1.9%; Score 39.2; DB 2; Length 4937;
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 803 GCTTATTATTGAAAACTACTCAAGTTGGATGTGAATGCATCCCGGCGAGGGTATTGAAGA 862
Db 4287 GGTATTATACCAACCAAGATGCTAAGCTAAATGGTGATGCATCAGGTGATGACAGAAGT 4346

QY 863 TGTCTGAAGAAAACAGCAACTCAAACTTGTGTGGGACAGATCCACGGAAGGATTGTTTAA 922
Db 4347 GAATGCAGTCAACGCAAGCGGCTCTGGTGTGAGTGTGACTCGGCGCAACCTCAAGCAGTGTGAA 4406

QY 923 TATGGATGAAGATGAAGAACTGAACATGAACAA 958
Db 4407 TATCACTGGGGATTAAACACAGTAAATGGGTAAA 4442

RESULT 13
US-08-728-470-3
; Sequence 3, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-728-470-3

Query Match
Best Local Similarity 1.9%; Score 39.2; DB 2; Length 4937;
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 803 GCTTATTATTGAAAACTACTCAAGTTGGATGTGAATGCATCCCGGCGAGGGTATTGAAGA 862
Db 4287 GGTATTATACCAACCAAGATGCTAAGCTAAATGGTGATGCATCAGGTGATGACAGAAGT 4346

QY 863 TGTCTGAAGAAAACAGCAACTCAAACTTGTGTGGGACAGATCCACGGAAGGATTGTTTAA 922
Db 4347 GAATGCAGTCAACGCAAGCGGCTCTGGTGTGAGTGTGACTCGGCGCAACCTCAAGCAGTGTGAA 4406

QY 923 TATGGATGAAGATGAAGAACTGAACATGAACAA 958
Db 4407 TATCACTGGGGATTAAACACAGTAAATGGGTAAA 4442

RESULT 14
US-08-617-697-3
; Sequence 3, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

US-08-617-697-3

Query Match 1.9%; Score 39.2; DB 2; Length 4937;  
Best Local Similarity 53.2%; Pred. No. 0.34;  
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 803 GCTTATTATTGAAACTACTCAAGTTGGATGTGAATGCATCCCGCAGGGTATTGAAGA 862  
Db 4287 GGTATTATTAACGCAAAAGATGCTTAAGCTAAATGGTGTGATGCATCAGGTGATGACAGAAGT 4346  
QY 863 TGCTGAAGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCCACGGAAGGATTGTTTAA 922  
Db 4347 GAATGCAGTCAACGCAAGCGGCTCTGTGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406  
QY 923 TATGATGAAGATGAAGAACTGAACATGAACAAA 958  
Db 4407 TATCACTGGGGATTTAAACACACAGTAATGGGTAAA 4442

RESULT 15

US-08-719-641-3  
; Sequence 3, Application US/08719641  
; Patent No. 6218141  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: Of No. 6218141-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/719,641  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-625  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4937 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-719-641-3

Query Match 1.9%; Score 39.2; DB 3; Length 4937;  
Best Local Similarity 53.2%; Pred. No. 0.34; Mismatches 0; Gaps 0;  
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 803 GCTTATTATTGAAACTACTCAAGTTGGATGTGAATGCATCCCGCAGGGTATTGAAGA 862  
Db 4287 GGTATTATTAACGCAAAAGATGCTTAAGCTAAATGGTGTGATGCATCAGGTGATGACAGAAGT 4346  
QY 863 TGCTGAAGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCCACGGAAGGATTGTTTAA 922  
Db 4347 GAATGCAGTCAACGCAAGCGGCTCTGTGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406  
QY 923 TATGATGAAGATGAAGAACTGAACATGAACAAA 958  
Db 4407 TATCACTGGGGATTTAAACACACAGTAATGGGTAAA 4442

Search completed: July 18, 2004, 06:25:34  
Job time : 156 secs